UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CAA-TP-86-2	2. GOVT ACCESSION NO. ADF-86-0122	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE): PROGRESS REPORT FOR THE PERIOD AUGUST 1984-JUNE 1985		5. TYPE OF REPORT & PERIOD COVERED Interim, Aug 84-June 85 6. PERFORMING ORG. REPORT NUMBER CAA-TP-86-2
7. AUTHOR(s) Dr. Robert L. Helmbold Dr. Aqeel A. Khan		8. CONTRACT OR GRANT NUMBER(s) N/A
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Concepts Analysis Agency 8120 Woodmont Avenue Bethesda, MD 20814-2797		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS (same)		12. REPORT DATE August 1986 13. NUMBER OF PAGES 271
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (same)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A
16. DISTRIBUTION STATEMENT (of this Report) Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) (same)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Joint Forces, Field Service, Operations, Statistical, System Analysis, Concept Development, Doctrinal Development, Wargame		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Describes work done on the Combat History Analysis Study Effort (CHASE) from August 1984 to June 1985. A data base on battles was coded into machine-readable form for analysis. The data were then summarized, examined for trends and interrelationships, and used to test hypotheses. The most appropriate topics for further investigation are also outlined.		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CAA-TP-86-2

COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE): PROGRESS REPORT FOR THE PERIOD AUGUST 1984 - JUNE 1985

August 1986

Prepared by

REQUIREMENTS AND RESOURCES DIRECTORATE

US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20814-2797

TECHNICAL PAPER
CAA-TP-86-2

ADF860122

COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE)

PROGRESS REPORT

FOR THE PERIOD AUGUST 1984 - JUNE 1985

AUGUST 1986

PREPARED BY
REQUIREMENTS AND RESOURCES DIRECTORATE
US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

DISCLAIMER

The findings of this report are not to be construed as an official Department of the Army position, policy, or decision unless so designated by other official documentation. Comments or suggestions should be addressed to:

Director
US Army Concepts Analysis Agency
ATTN: CSCA-RQ
8120 Woodmont Avenue
Bethesda, MD 20814-2797

DEPARTMENT OF THE ARMY
US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

REPLY TO
ATTENTION OF:

CSCA-SAMV

11 MAR 1987

SUBJECT: Combat History Analysis Study Effort

Deputy Under Secretary of the Army (OR)
Headquarters, Department of the Army
Washington, DC 20310

1. U.S. Army Concepts Analysis Agency (CAA) initiated the Combat History Analysis Study Effort (CHASE) in August 1984 to search for historically-based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses.

2. Progress made in the period August 1984-June 1985 is reported in the enclosed Technical Paper. It indicates that data on historical battles can be used to discover quantitative trends and relations of potential significance to military operations research, concept formulation, wargaming, and studies and analyses.

3. At the same time, additional research is needed to pursue the new lines of investigation suggested by this initial effort, and to clarify some of the anomalies it has turned up.

4. Despite its tentative and unfinished state, the work described in this Technical Paper is being provided to you now in the expectation that those interested in the scientific and quantitative aspects of military operations research will find it beneficial to their efforts. Questions or inquiries should be directed to the Special Assistant for Models Validation, U.S. Army Concepts Analysis Agency (CSCA-SAMV), 8120 Woodmont Avenue, Bethesda, MD 20814-2797, (301) 295-1669.

E. B. Vandiver III

Encl

E. B. VANDIVER III
Director

COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE): PROGRESS REPORT FOR THE PERIOD AUGUST 1984 - JUNE 1985	STUDY SUMMARY CAA-TP-86-2

THE REASON FOR PERFORMING THE STUDY was to carry out the initial phase of the Combat History Analysis Study Effort (CHASE), whose ultimate purpose is to search for historically-based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses.

THE PRINCIPAL FINDING of the work done during the period covered by this paper (August 1984 to June 1985) is that data on historical battles can be used to discover quantitative trends and relations of potential significance to military operations research, concept formulation, wargaming, and studies and analyses.

THE MAIN ASSUMPTIONS on which the CHASE Study, as well as its major phases, rests are:

(1) Historical battle data can be analyzed using modern statistical methods.

(2) Formulas are not to be complicated without good empirical evidence.

(3) Long-term trends and relations can be extrapolated to future situations with a reasonable degree of confidence.

THE PRINCIPAL LIMITATIONS which may affect the findings presented in this progress report are as follows:

(1) Data on strengths at intermediate stages during the course of a battle were not available for use in this phase of the CHASE Study.

(2) The study used a data base prepared for the US Army Concepts Analysis Agency (CAA) by the Historical and Research Evaluation Organization (HERO). The HERO data base, even though composed of 601 battles, is still not large enough to support adequately all of the statistical analyses that should be performed.

(3) Typographical mistakes, omissions, ambiguities and ill-defined data categories in the HERO data base weakened some of the analysis results, and precluded some analyses that would have been desirable.

(4) Because of data inadequacies and the limited scope of this initial phase of the CHASE Study, not all of CHASE's Essential Elements of Analysis (EEAs) could be fully addressed.

THE SCOPE OF THE WORK done during the period covered by this progress report, was limited to an initial analysis of the HERO data base of 601 battles. This scope included:

(1) Reducing to machine-readable form all of the tabulated data in the HERO data base.

(2) Assessing the suitability of the data base for quantitative analysis.

(3) Summarizing selected portions of these data to facilitate their efficient use in military operations research, concept formulation, wargaming, and studies and analyses.

(4) Seeking important trends and interrelations present but hidden in these data.

(5) Testing selected hypotheses against the data.

THE STUDY OBJECTIVE for the period covered by this progress report included:

(1) Evaluating the suitability of the HERO data base for quantitative analysis, identifying essential data base improvements, and taking necessary corrective measures.

(2) Experimenting with a variety of analytical techniques to assess their ability to expose quantitative trends and relations of significant potential use in military operations reserch, concept formulation, wargaming, and studies and analyses.

(3) Identifying specific issues for further investigation in subsequent phases of the CHASE Study.

THE STUDY SPONSOR was the US Army Concepts Analysis Agency.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Resources and Requirements Directorate.

COMMENTS AND SUGGESTIONS may be sent to the Director, US Army Concepts Analysis Agency, ATTN: CSCA-RQ, 8120 Woodmont Avenue, Bethesda, MD 20814-2797.

Tear-out copies of this synopsis are at back cover.

CONTENTS

FIGURES

FIGURE		Page

TABLES

TABLES

PREFACE

This paper documents the work done on the Combat History Analysis Study Effort (CHASE) during the period August 1984 - June 1985. This progress report is presented as a standalone document with the expectation that those interested in the scientific and quantitative aspects of military operations research, concept formulation, wargaming, and studies and analyses will find it beneficial to their efforts.

However, readers are cautioned that this paper is an interim progress report of continuing research, intended in the first instance for internal use at the US Army Concepts Analysis Agency (CAA). Subsequent phases of CHASE should improve on it, and as our insight deepens many of its findings and observations may require substantial modification.

Some readers may find this paper hard to read. It has been said* that "It is so customary for political writing to flow on with journalistic ease that people seem to regard ease as a characteristic of thought about politics, whereas really it is only a characteristic of popularization." Although this paper may not "flow on with journalistic ease," we hope that its scientific approach to combat dynamics will interest readers enough to make its study worthwhile.

There are some who might object that a science of combat dynamics is impossible because combat is so strongly influenced by the actions of people. Marshall** has made the following eloquent remarks applicable to that objection: "The actions of men are so various and uncertain, that the best statement of tendencies, which we can make in a science of human conduct, must needs be inexact and faulty. This might be urged as a reason against making any statements at all on the subject; but that would be almost to abandon life. Life is human conduct, and the thoughts and emotions that grow up around it. By the fundamental impulses of our nature we all--high and low, learned and unlearned--are in our several degrees constantly striving to understand the courses of human action, and to shape them for our purposes, whether selfish or unselfish, whether noble or ignoble. And since we must form to ourselves some notions of the tendencies of human action, our choice is between forming those notions carelessly and forming them carefully. The harder the task, the greater the need for steady patient inquiry; for turning to account the experience that has been reaped by the more advanced physical sciences; and for framing as best we can well thought-out estimates, or provisional laws, of the tendencies of human action." The work described in this paper is offered in this spirit.

---

*Richardson, Lewis Fry, "Statistics of Deadly Quarrels," The Boxwood Press, Pacific Grove, CA, 1960.

**Marshall, Alfred, "Principles of Economics," 1890.

COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE):

PROGRESS REPORT FOR THE PERIOD AUGUST 1984 - JUNE 1985

CHAPTER 1

EXECUTIVE SUMMARY

1-1. PROBLEM. Although the works on military history are of considerable interest and utility to practitioners of the military art, few of them are in a form suitable for direct application to military operations research, concept formulation, wargaming, asd studies and analyses. Usually, these activities can use efficiently only such historical combat experience that is expressed in the form of mathematically explicit quantitative relations that are universally applicable throughout an extremely wide range of engagement situations. The Combat History Analysis Study Effort (CHASE) was established to search for historically based quantitative results that are suitable for use in military operations research, concept formulation, wargaming, and studies and analyses.

1-2. BACKGROUND. In 1983 and 1984, the Historical Evaluation and Research Organization (HERO) prepared for the US Army Concepts Analysis Agency (CAA), under Contract No. MDA903-82-C-0363, a data base of 601 battles and engagements. This was published in 1984 **(Ref 1-1)**, and will be referred to as the HERO data base. As that effort was drawing to a close it was realized that, although the HERO data base is unique and of great potential value because it is detailed for individual battles, it is not directly usable in CAA studies and analyses because it does not explicity provide quantitative trends and interrelations. As a result, CAA established the CHASE project, with the objective of searching for historically based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses.

1-3. SCOPE

a. The overall scope of the CHASE Study includes the following:

(1) Reduce all or a significant portion of the HERO data base to machine-readable form for analysis.

(2) Summarize the mass of data in the HERO data base and present the results for use in military operations research, concept formulation, wargaming, and studies and analyses.

(3) Seek trends and interrelationships present but hidden in the data.

(4) Test selected hypotheses against the data.

b. This paper documents the progress made on the CHASE Study in its initial phase (August 1984 - June 1985). The scope of the effort undertaken during this period included the following:

(1) Reduce to machine-readable form all of the tabular data in the HERO data base. The result of this step will be referred to as the computerized data base.

(2) Proofread and review for accuracy and consistency the data presented in the HERO data base. This led in a natural way to the establishment of a new contract with HERO to eliminate some of the typographical mistakes, omissions, inconsistencies, ambiguities, and redundancies discovered in the HERO data base, and to expand it in selected areas.

(3) Explore the prospects for using these data to obtain quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses. This included preparing (or locating) computer programs suitable for manipulating and analyzing the computerized data base, and then applying them appropriately to create selected descriptive or summary statistical tabulations of the data, to seek factors associated with victory in battle, to test selected hypotheses against the data, and to explore ways to reduce some of the redundancies present in the data.

(4) Plan the most important next steps for accomplishing the CHASE Study in light of the experience gained to date.

1-4. LIMITATIONS. The principal limitations which may affect the findings presented in this progress report are as follows:

a. Data on strengths at intermediate stages during the course of a battle were not available for use in this phase of the CHASE Study.

b. The study used almost exclusively the HERO data base which, even though composed of 601 battles, is still not large enough to support adequately all of the statistical analyses that should be performed.

c. Typographical mistakes, omissions, ambiguities and ill-defined data categories in the HERO data base weakened some of the analysis results, and precluded some analyses that would have been desirable.

d. Because of data inadequacies and the limited scope of the initial phase of the analysis, not all of CHASE's Essential Elements of Analysis (EEAs) were fully addressed. Subsequent phases of the CHASE Study will fill these voids.

1-5. TIMEFRAME. The computerized data base contains information on 601 battles that took place between 1600 and 1973. In a few places, data on battles from earlier times are used to supplement the computerized data base. This paper presents findings only for those trends or relations that have persisted relatively unchanged over long periods of time, and which thus appear to be extrapolatable to future situations with a reasonable degree of confidence.

1-6. KEY ASSUMPTIONS. The main assumptions on which the CHASE Study, as well as its major phases, rests are:

a. Historical battle data can be analyzed using modern statistical methods.

b. Formulas are not to be complicated without good empirical evidence.

c. Long-term trends and relations can be extrapolated to future situations with a reasonable degree of confidence.

1-7. APPROACH. The approach adopted during the period covered by this paper is as follows:

a. A data base format for use in computerizing the HERO data base was designed. The tabular data in the HERO data base were then computerized using that data base format. As data were transcribed into the computerized data base, a written record was kept of any missing, confusing, or questionable data items in the HERO data base. The computerized data were manually proofread against the HERO data base twice--once immediately after each table from the HERO data base was entered into the computerized data base, and again after the computerized data base had been completed. In addition, a computer program was written to check that each entry in the computerized data base is within its legitimate range. This computer program also made some selected checks on the consistency of the HERO data. For example, it checked to see that attacker and defender achievement ratings were consistent with the designation of victorious side. All differences between the computerized and the HERO data bases discovered by these manual and automated checks were corrected before the computerized data base was analyzed.

b. The subsequent analysis of the computerized data began with an informal examination and some simple summarizations of the data (descriptive statistics). It then progressed to searching for the factors associated with victory. Because it was determined that some of the data in the HERO data base were at least partially redundant, factor analysis techniques were explored to assist in understanding this redundancy. Finally, a test of a particular hypothesis regarding breakpoints was carried out.

c. Throughout all stages of the study a determined effort was made to apply to the analysis of these data on historical battles the most powerful and appropriate modern statistical techniques and data processing technologies. It was, of course, necessary to tailor the analytical approach to the particular issue being investigated, and in fact a wide variety of techniques were employed in one part of the study or another. The most frequently used techniques employed in the period covered by this report include:

- Graphical and exploratory data analysis techniques such as scatter diagrams.
- Construction of histograms and empirical distribution functions.
- Contingency table analysis.
- Curve and function fitting methods such as linear and logistic regression.
- Correlation and factor analysis.

d. Wherever possible, an attempt was made to follow the precepts of the method of "strong inference" **(Ref 1-2)** and the method of "multiple working hypotheses" **(Ref 1-3)**. These methods involve the systematic consideration of well-defined alternative hypotheses, the deduction from these hypotheses of consequences that are testable against the available data, the design of crucial experiments that will discriminate sharply against one or more of the alternative hypotheses, and the deliberate search for important new hypotheses. Consequently, new areas for future investigation are identified and documented.

e. A conscientious attempt is made to adhere to high standards of scientific investigation. Very little is assumed about the structure or dynamics of combat. Instead, the guiding principle is that a hypothesis or widely held opinion regarding battle is not to be taken for granted, but that the data are to be consulted to determine whether they support it or not. Therefore, frequent (though usually implicit) appeal is made to various forms of the well-known principle of Ockham's Razor to the effect that "Entities are not to be multiplied without necessity" **(Ref 1-4)**. The following versions of this principle are frequently used to focus inquiry on substantive issues:

(1) "Formulae are not to be complicated without good evidence." **(Ref. 1-5)**.

(2) "Complications in models are not to be multiplied beyond the necessity of practical application and insight" **(Ref 1-6)**.

(3) "The burden of proof is on the party claiming that such-and-such a factor must be introduced to explain the data. The claimant must show that the data are incompatible with the simpler theory in which the new factor is left out, but that they are compatible with the more complicated theory that arises when the new factor is introduced" **(Ref 1-7)**.

(4) "A hypothesis that cannot be confronted with hard evidence is metaphysical, and may safely be ignored" **(Ref 1-8)**.

f. It might be thought that the methods used presume the existence of patterns in history that can be discovered. But it would be more correct to say that the existence of such patterns is itself a hypothesis that can be tested by searching for them. If some patterns are found, then they exist. If, after sufficiently diligent search, no patterns are found, then this constitutes evidence for the hypothesis that no such patterns exist--just as the search for perpetual motion machines led ultimately to the hypothesis that no such machines are physically realizable.

1-8. ESSENTIAL ELEMENTS OF ANALYSIS AND ANSWERS. The research was guided by five EEAs, as provided by the Study Directive (Appendix B). Summaries of the state of development reached during the period covered by this paper are given below.

a. **Can the Factors Associated with Victory in Battle be Identified?** Six variables were tested for close association with victory in battle. Each of the variables is an explicit, mathematically defined function of the tabulated data on personnel strengths and losses. (Chapter 4, provides a full technical definition of these variables, and the Glossary contains summary definitions of them.) The six variables included the force ratio (FR), the casualty exhange ratio (CER), the fractional exchange ratio (FER), a measure of the bitterness of a battle (or total losses to both sides) (EPS), a theoretically-motivated index of the defender's advantage vis-a-vis the attacker (ADV), and a measure of the residual portion of ADV after the average effect of force ratio on it has been removed (RESADV). Of these six variables, the defender's advantage (ADV) and the fractional exchange ratio (FER) are most closely associated with victory in battle. RESADV and CER are somewhat less closely associated with victory in battle. EPS and FR are substantially less closely associated with victory in battle. Some of the battles of the World War II (and some later) eras seem to be anamalous in the sense that for these battles the relationship of victory in battle to ADV is much weaker than for battles of other eras, and for most other battles of the same era. The reasons why these battles are anomalous, and why they more prevalent during the WW II and later eras, is not yet well understood. However, the leading hypothesis at the moment appears to be that the data for several battles of the WW II and later eras is flawed.

b. **What long-term trends can be detected in historical combat data?** The analysis of long-term trends was not emphasized during the period covered by this paper. However, it appears that there has been no long-term secular trend over the last 400 years in the proportion of battles won by the attacker.

c. **Can the historical influence of air support on the outcome of land battles be quantified?** An analysis of the effects of air support was not within the scope of the effort covered by this paper.

d. **What can be said about the factors influencing rates of advance in land combat?** An analysis of the factors influencing rates of advance was not considered fruitful during the period covered by this paper, because the battle duration data in the data base used were reported only to the nearest day, which is too coarse a time resolution to provide rate values suitable for analysis.

e. **What lessons were learned regarding the preparation of battle and engagement data bases for use in quantitative analyses?** Lessons learned regarding the preparation of data bases will be reported separately, in accordance with the study plan.

1-9. OTHER KEY FINDINGS

a. The HERO data base needs to be enhanced before analyzing it extensively. To satisfy the need for data base refinement, a contract was awarded to the Historical Evaluation and Research Organization (HERO) to revise and extend the data base. The results of this contract were not available during the period of time covered by this paper (August 1984-June 1985).

b. The data base is mainly typical of organized division- to corps-level forces engaged in intense, short (hours to days) battles in Europe and America during the nineteenth and twentieth centuries.

c. Battle durations seem to fit a Weibull or a lognormal distribution about equally well.

d. Casualty fractions seem to be distributed approximately lognormally. The attacker's casualty fraction tends to be less than the defender's.

e. The personnel force ratio (FR), personnel casualty exchange ratio (CER), and the personnel fractional exchange ratio (FER) are all approximately lognormally distributed.

f. Force ratio is an unsatisfactory and inadequate predictor of victory in battle. Both advantage (ADV) and fractional exchange ratio (FER) (see the Glossary at the end of this paper) are much more closely related to victory than is the force ratio. Consequently, either advantage or fractional exchange ratio should be used as a figure of merit for comparing force structures, contingency plans, equipment options, and tactics in simulation experiments.

g. There is a high degree of redundancy among some of the items in the data base. The analysis of this redundancy, and the development of measures to deal correctly and effectively with it, need further investigation.

h. When a breakpoint hypothesis similar to those conventionally used to terminate simulations and wargames is tested against the HERO data base, it is found to be inconsistent with the data. The reasons for this are not yet well understood.

CHAPTER 2

SOURCES OF DATA ON BATTLES AND ENGAGEMENTS

2-1. INTRODUCTION. This chapter describes the data base used as the source of data on battles and engagements throughout the period covered by this progress report, presents the design and implementation of the computerized data base, indicates some of the problem areas uncovered in this process, introduces some terminology that will be used throughout subsequent portions of this paper, and cites some other data bases that may be found useful in future work.

2-2. THE HERO DATA BASE

a. In 1984, the US Army Concepts Analysis Agency (CAA) published the HERO data base of battles and engagements **(Ref 2-1)**. This data base provides detailed data on each of 601 battles from the period 1600 AD to 1973 AD. The distribution of battle dates over time, along with some other descriptive statistics of the material in the HERO data base, is discussed in Chapter 3. The HERO data base consists of seven tables covering:

(1) Battle identification (name, dates, campaign, war, forces and commanders involved, duration, and width of front).

(2) Operational and environmental variables (defender posture, terrain, weather, season, surprise, air superiority).

(3) Strengths and losses on both sides.

(4) Intangible factors (such as combat effectiveness, leadership, training, etc.).

(5) Outcome (victorious side, distance advanced, mission accomplishment of each side).

(6) Factors affecting the outcome (such as force quality, reserves, air superiority, etc.).

(7) Combat forms and resolution of combat (main attack and scheme of defense, secondary attack, resolution of the combat).

Tables 2-1 through 2-6 give a sample of the kinds of data presented in the HERO data base tables. Appendix E gives an extended description of the information included in each HERO data base table. In all, almost 90 items of information are tabulated for each of the 601 battles in this data base. The HERO data base values recorded in Tables 1 and 3 are objective quantities that, at least in principle, all observers could agree upon if completely trustworthy reports were available. The values recorded in HERO's Tables 2, 5, and 7, however, are overall impressions and more difficult to objectify in a manner acceptable to all observers, even if completely trustworthy reports were available. The values recorded in HERO's Tables 4 and 6 are frankly judgmental, and hence almost impossible to objectify in a manner acceptable to all observers. The reader is referred to CAA-SR-84-6 **(Ref 2-1)** for a complete picture of the HERO data base.

Table 2-1. Example of HERO Data Base (Table 1)

Engagement		Date(s)	Campaign	Forces	Commanders	Duration (days)	Width of Front (Km)
Murfreesboro, Tennessee	A D	31 Dec 1862- 3 Jan 1863	Stones River	CS Army of Tennessee US Army of the Cumb'd	Bragg Rosecrans	4	7.0
Chancellorsville, Virginia	A D	1-6 May 1863	Chancellorsville	US Army of the Potomac CS Army of No. Va.	Hooker Lee	6	25.0
Champion's Hill, Mississippi	A D	16 May 1863	Vicksburg	US Army CS Army	Grant Pemberton	1	6.4
Brandy Station, Virginia	A D	9 Jun 1863	Gettysburg	US Cav. Corps CS Cav. Corps	Pleasanton Stuart	1	8.0
Gettysburg, Pennsylvania	A D	1-3 Jul 1863	Gettysburg	CS Army of No. Va. US Army of the Potomac	Lee Meade	3	10.5
Chickamauga, Georgia	A D	19-20 Sep 1863	Chickamauga	CS Army of Tennessee US Army of the Cumb'd	Bragg Rosecrans	2	10.0
Chattanooga, Tennessee	A	24-25 Nov 1863	Chattanooga	US Army of the Cumb'd CS Army of Tennessee	Grant Bragg	2	16.0

Table 2-2. Example of HERO Data Base (Table 2)

Engagement		Defender Posture	Terrain	Weather	Season	Surprise	Surpriser	Level Surprise
Murfreesboro	A D	HD	RM	WLC	WT	Y	x	Substantial
Chancellorsville	A D	HD	RM	DST	SpT	Y	x	Complete
Champion's Hill	A D	HD	RM	DST	SpT	N	--	--
Brandy Station	A D	HD	RM	DST	ST	Y	x	Substantial
Gettysburg	A D	HD	RM	DST	ST	N	--	--
Chickamauga	A D	HD	RM	DST	FT	Y	x	Substantial
Chattanooga	A D	P /FD	RgM, RM	WL/DST	FT	N	--	--

Table 2-3. Example of HERO Data Base (Table 3)

Engagement		Strength			Battle Casualties		Arty. Pieces Lost		Success	Advance (Km/Day)
		Total	Cavalry	Arty. Pieces	Total	%/Day	Total	%/Day		
Murfreesboro	A	34,732	4,500	120	11,739	8.4	6	1.3	x	2.0
	D	41,400	3,200	100	12,906	7.8	28	7.0	x	--
Chancellorsville	A	134,000	?	404	17,278	2.1	120	5.0		0
	D	80,000	?	170	12,821	2.7	7	0.7	x	--
Champion's Hill	A	29,373	?	?	2,441	8.3	?	--	x	2.0
	D	20,000	500	?	3,851	19.3	11	--		--
Brandy Station	A	12,000	?	?	900	7.5	?	--	x	1.5
	D	10,000	?	?	500	5.0	?	--		--
Gettysburg	A	75,054	8,000	250	28,063	12.5	3	0.4		1.1
	D	83,289	13,000	300	23,049	9.2	6	0.7	x	--
Chickamauga	A	66,326	8,000	?	18,454	13.9	15	--	x	1.6
	D	58,222	10,000	246	16,170	13.9	51	10.4		--
Chattanooga	A	61,000	?	?	5,824	4.8	?	--	x	4.4
	D	40,000	4,856	?	6,667	8.3	40	--		--

Table 2-4. Example of HERO Data Base (Tables 4 and 5)

Engagement		CE	Leader-ship	Training/ Experience	Morale	Logis-tics	Momen-tum	Intelli-gence	Tech-nology	Initia-tive	Victor	Distance Advanced (Km/Day)	Mission Accomp.
Murfreesboro	A	C	C	C	C	N	N	N	C	x	x	2.0	6
	D										x	--	5
Chancellors-ville	A	C		C	C	N	N		C			0	3
	D		x					x		x	x	--	10
Champion's Hill	A	C	x	C	C	N	N	N	C	x	x	2.0	8
	D											--	4
Brandy Station	A	C	C	C	C	N	N	x	C	x	x	1.5	6
	D											--	5
Gettysburg	A	C	C	C	C	N	N		C	x		1.1	4
	D							x			x	--	6
Chickamauga	A	C		C	C	N	N	x	C	x	x	1.6	6
	D		0									--	4
Chattanooga	A	C	x	C	C	N	N	N	C	x	x	4.4	8
	D											--	4

Table 2-5. Example of HERO Data Base (Table 6)

Engagement		Force Quality	Reserves	Mobility Superiority	Force Preponderance	Weather	Terrain, Roads	Leadership	Planning	Surprise	Maneuver, Mass, Narrow front	Logistics	Fortifications	Depth
Murfreesboro	A	N	N	N	N	N	N	N	N	x	N	N	N	N
	D													
Chancellorsville	A	N	N	N	N	N	N				N	N	N	N
	D							x	x	x				
Champion's Hill	A	N	N	N	N	N		x	N	N	N	N	N	N
	D						x							
Brandy Station	A	N	N	N	N	N	N	N	x	x	N	N	N	N
	D													
Gettysburg	A	N		N		N		N	N	N	O	N	N	N
	D		x		x		x							
Chickamauga	A	N	N	N	N	N			x	x	x	N	N	N
	D						x	O						
Chattanooga	A	N	N	N	N	N		x	x	N	N	N		N
							x						x	

Table 2-6. Example of HERO Data Base (Table 7)

Engagement		Plan and Maneuver: Main Attack and Scheme of Defense	Plan and Maneuver: Secondary Attack	Success	Resolution
Murfreesboro	A	F, EE	--		P, S, WD
	D	D	--	x	
Chancellorsville	A	E(LR)	F(RF)		R, WD
	D	D/O, E(RR)	--	x	B
Champion's Hill	A	F	--	x	P, Ps
	D	D	--		WD
Brandy Station	A	F, E(RR)	--	x	P, WD
	D	D/O	--		--
Gettysburg	A	F, EE	--		R, WD
	D	D	--	x	--
Chickamauga	A	F	--	x	P, Ps
	D	D	--		WD
Chattanooga	A	F, EE	F, P	x	B, Ps
	D	D	--		WD

2-3. THE COMPUTERIZED VERSION OF THE HERO DATA BASE

a. In order to facilitate the manipulation and analysis of these data, they were encoded in computer-readable data files. Appendix F describes the coding scheme used for this purpose. Table 2-7 provides a sample of the computerized version of the HERO data base. The specific data file formats for the computerized data base are given in Appendix G. Appendix H provides an index of the battles and engagements in the computerized data base.

Table 2-7. Sample Entry from CAA Computerized Data Base

```
NO. =    199  NAME = GETTYSBURG                                    LOCN = PENNSYLVANIA
WAR = AMERICAN CIVIL WAR                              CAMPGN = GETTYSBURG
DATE =  1 JUL 1863  1 =      3  WOF =        10.5
NAMA=CS ARMY OF NORTHERN VIRGINIA                                   COA=LEE
NAMD=US ARMY OF THE POTOMAC                                         COD=MEADE

POSTD1 = HD      POSTD2 = OO
TERRA1 = RMD     TERRA2 = COD
WX1 = DSTST         WX2 = 00000      WX3 = 00000
SURPA =       .9  AEROA =     0

        XO/YO    CX/CY      CAV    TANK     LT     MBT     ARTY     FLY    CTANK   CARTY    CFLY
ATT     75054    28063     8060       0      0       0      250       0        0       3       0
DEF     83289    23049    13000       0      0       0      300       0        0       6       0

     CEA   LEADA   TRNGA  MORALA   LOGSA  MOMNTA  INTELA   TECHA   INITA
       0       0       0       0       0       0      -1       0       1

 WINA     KPDA ACHA ACHD
   -1      1.1    4    6

    QUALA     RESA  MOBILA    AIRA  FPREPA    WXA   TERRA  LEADAA   PLANA  SURPAA    MANA  LOGSAA  FORTSA   DLEPA
        0       -1       0       0      -1      0      -1       0       0       0      -1       0       0       0

      PRI1  PRI2  PRI3  SEC1  SEC2  SEC3  RESO1 RESO2 RESO3
ATT     FF    DE    OO    OO    OO    OO    RR    WD    OO
DEF     DD    OO    OO    OO    OO    OO    OO    OO    OO
WGT = MED
```

b. While the computerized data base was being prepared, written records were kept of missing, confusing, ambiguous, or questionable data items. Slightly over 400 of these "Data Base Problem Reports" were eventually accumulated documenting omissions, inconsistencies, ambiguities, redundancies, and typographical errors in the HERO data base. Table 2-8 gives a few examples of the kind of problems that were surfaced in this manner. Table 2-9 lists the battles for which at least one of the XO, YO, CX, or CY values was missing. Here, and throughout the rest of this paper, we use the symbols XO and YO for the attacker's and defender's (respectively) total personnel strength, and CX and CY for the attacker's and defender's (respectively) personnel losses. We also use ATK and DEF as abbreviations for attacker and defender. Table 2-9 lists for illustration some of the missing data items in the HERO data base. These 16 battles have to be omitted from all tabulations involving casualties or losses, and three of them have to be omitted from tabulations involving force strengths.

Table 2-8. Data Base Problem Reports

- About 400+ problems noted
- Examples:

  1. Buzancy Ridge, ATK force = US 18th Inf Rgt (-) (+) (but the previous battle, on the same date, with same commanding officer, gives ATK force = US 28th Inf Rgt (-) (+))

  2. Iwo Jima (final phase), ATK strength = 32,000, DEF strength = 2,685, width of front = 1.8 km (but can the ATK force engage all of its troops under these conditions?)

  3. Egyptian offensive north, ATK withdrew with heavy losses (but ATK losses were only 2.1%)

  4. Brusilov offensive, ATK stalemated (but was rated 7 out of 10 for achievement and credited with winning, while DEF withdrew with heavy losses)

Table 2-9. HERO Data Base Battles Having Missing Personnel Strength or Casualty Data

No.	ISEQNO	Name	Missing data items
1	26	Preston	CX
2	40	Killiecrankie	CX, CY
3	216	Dinwiddie Courthouse	CY
4	248	Kumanovo	CY
5	254	The Nieman	CX
6	267	Le Cateau	CX
7	289	Eastern Champagne[a]	XO, YO
8	291	Ypres II[a]	XO, YO
9	292	Festubert[a]	YO
10	300	First Dardanelles landing	CY
11	301	Suvla Bay	CY
12	391	Chouigi Pass	CX, CY
13	461	Mortain	CX
14	469	Schmidt	CY
15	484	St. Vith	CX, CY
16	485	Bastogne	CX, CY

[a]Missing XO, YO, or both.

c. These problems indicated a need to enhance the HERO data base before analyzing it extensively. To satisfy this need, a contract was awarded to the Historical Evaluation and Research Organization (HERO) to revise and extend the work presented in CAA-SR-84-6 **(Ref 2-1)**. This contract will be referred to as the CHASE Data Enhancement Study (CDES). The CDES contract was awarded, and work on it was begun on 6 June 1985, with an anticipated completion date of December 1985 (subsequently extended to January 1986). It calls for accomplishment of the nine tasks enumerated in Table 2-10 and further detailed in Appendix I. No results from the CDES contract are included in this progress report, which covers only the period August 1984 through June 1985.

Table 2-10. CHASE Data Enhancement Study (CDES) Contract Tasks

1. Analyze the data base problem reports.
2. Clarify the total engaged personnel strength.
3. Clarify the basis for assigning victory.
4. Refine the duration data.
5. Clarify the width of front data.
6. Clarify the defender posture description.
7. Identify the quality of strength and loss data.
8. Develop strength and attrition histories for selected battles.
9. Assist in eliminating unwanted redundancies.

2-4. ADDITIONAL DATA BASES

a. Several other data bases were considered for use in the CHASE Study. As shown in Table 2-11, the three most important data bases of land combat battles and engagements for our purposes are the HERO data base (see paragraph 2-2, above), the Combat Operations Research Group (CORG) data base described in several CORG reports **(Refs 2-2 through 2-4)**, and the Bodart-Willard-Schmieman (BWS) data base **(Ref 2-5)**. The latter originated with Bodart's Kriegslexicon **(Ref 2-6)**, which was computerized by Willard **(Ref 2-7)**, and subsequently modified by Schmieman **(Ref 2-8)**. These three major data bases overlap in the sense that some battles (e.g., Gettysburg) are listed in two or more of them. As indicated in Table 2-11, some important supplemental information on the battles and engagements contained in the three major data bases is provided in several books **(Ref 2-6, 2-9, 2-10, 2-11, and 2-12)**. However, there are hardly any battles in these supplemental references that are not already in at least one of the three major data bases listed in Table 2-11.

Table 2-11. Major Data Bases of Land Combat Battles and Engagements

Data base	Number of battles	Dates covered	Date appeared
HERO/CAA	601	1600-1973	1983-84
CORG	175	280 BC-1945 AD	1961-63
Bodart-Willard-Schmieman (BWS)	ca. 1,000[a]	1618-1905	1908-67
Key supplemental information:			
HERO QJM Data Base (book)	204	1943-1973	1979
Livermore "Numbers & Losses"	64	1861-1865	1900
Dodge "Napoleon"	100	1631-1815	1907
Bodart "Kriegslexicon"	ca. 1,000[a]	1618-1905	1908
Berndt "Zahl Im Kriege"	91[b]	1741-1871	1897

[a]Of which ca. 100 are sieges.
[b]Of which 13 are sieges.

b. In the period covered by this progress report, only the HERO data base was used. In future phases of the CHASE project, the other major data bases (CORG and BWS) can be used to extend, refine, or confirm the major findings obtained by using the HERO data base. Some additional effort will be required to put those data bases in a form suitable for such use.

2-5. SUBSAMPLES. It is sometimes desirable to extract from the data base selected subsamples, which are used for specific purposes.

a. One of the subsamples used during the period covered by this progress report, called the exploratory subsample, consists of a random sample of 100 battles taken from the HERO data base battles with starting dates earlier than 1943. It was used for some of the exploratory statistical work, especially in the data redundancy analysis described in Chapter 5. It was also used to develop, test and debug many of the statistical analysis procedures and computer programs used to examine the computerized data base. Table 2-12 lists the sequence numbers of the battles included in the exploratory subsample.

Table 2-12. List of Battles Included in the Exploratory Subsample

Exploratory subsample number	ISEQNO	Exploratory subsample number	ISEQNO
1	6	51	206
2	7	52	214
3	9	53	229
4	10	54	230
5	13	55	232
6	18	56	235
7	23	57	246
8	26	58	252
9	30	59	254
10	34	60	261
11	43	61	265
12	44	62	267
13	46	63	271
14	48	64	275
15	50	65	282
16	61	66	283
17	69	67	287
18	74	68	290
19	76	69	292
20	80	70	294
21	81	71	296
22	88	72	303
23	91	73	306
24	92	74	311
25	97	75	312
26	100	76	315
27	103	77	322
28	104	78	326
29	114	79	327
30	115	80	329
31	119	81	330
32	124	82	332
33	126	83	333
34	131	84	336
35	132	85	338
36	136	86	342
37	144	87	343
38	150	88	345
39	151	89	347
40	157	90	348
41	160	91	353
42	166	92	357
43	171	93	359
44	183	94	363
45	189	95	365
46	191	96	373
47	200	97	375
48	201	98	381
49	202	99	387
50	203	100	389

b. Other subsamples used are the WWII and the non-WWII subsamples. The WWII subsample consists of the 163 battles in the computerized data base that started between 19400101 and 19491231. Here dates are given in the form YYYYMMDD with YYYY indicating the year, MM the number of the month, and DD the number of the day. For example, 19400101 means that the year is 1940, the month is number 1 (January), and the day is 1. The non-WWII subsample consists of all battles in the computerized data base other than those in the WWII subsample. Additional subsamples are defined as needed in subsequent chapters.

2-6. NEXT STEPS REGARDING DATA BASES. The anticipated next steps for data base work in support of CHASE include the items listed in Table 2-13.

Table 2-13. Next Steps for Data Bases

- Accomplish CDES contract tasks 1-9
- Revise and extend the computerized data base accordingly
- Purge the data base of all additional known or suspected errors
- Bring the BWS and CORG data bases on line
- Document the descriptions of these computerized data bases
- Document the lessons learned regarding the preparation of data bases on battles for use in quantitative analysis

2-7. CONCLUDING OBSERVATIONS ON DATA BASES

a. The HERO data base of 601 battles provides more detailed and systematically tabulated information on more battles, especially recent battles, than any other currently available data base. As a result, it often is better suited to quantitative analysis than other sources of information. The CDES contract results will substantially enhance its accuracy and utility.

b. Additional, less comprehensive data bases will usefully supplement information in the HERO data base, and can be used selectively to investigate the extent to which findings based on the HERO data extend to other data bases.

CHAPTER 3

DESCRIPTIVE STATISTICS

3-1. INTRODUCTION. This chapter presents some of the descriptive statistics generated using the computerized data base described in Chapter 2. Descriptive statistics merely express compactly the most salient features of the data, using the least sophisticated analysis techniques. This often makes them the easiest to understand. Consequently, it is important to see how much can be done with descriptive statistics, even though they are not usually powerful enough to cope with some of the deeper and potentially more important issues.

3-2. THE HERO DATA BASE REPRESENTS A WIDE RANGE OF COMBAT EXPERIENCE

a. Table 3-1 shows some general facts about the computerized data base. Note that the range of battle dates includes the colonization of Jamestown (1607) and Plymouth (1620), and the first safe visit of man to the moon (1969). The total engaged troop strength, obtained by summing the number of attacker and defender total strengths for all battles, amounts to about the population of Bangladesh, the eighth most populous nation on earth. The total battle casualties, obtained by summing the attacker's and defender's personnel casualties for all battles, is about equal to the population of New York state. The total battle days, obtained by summing the battle durations of all battles, amounts to about 6.3 years. The total distance advanced by the attacker (ATK), obtained by summing the distances advanced in individual battles, is about equal to the round-trip airline distance from Los Angeles, California to Pittsburgh, Pennsylvania. The total area gained by the attacker, obtained by summing the products of width of front and distance advanced for the individual battles, is about equal to the area of Peru, the nation with the nineteenth largest area. Clearly, an immense amount of battle experience is captured by this data base. The period of time covered spans an extremely broad range of technologies, and hence should allow important findings regarding trends to be derived.

b. However, it is also true that the computerized data base is mainly representative of short, pitched land combat battles fought by organized division- and corps-sized military formations during the 19th and early 20th Centuries in Europe or North America. The computerized data base contains no sea or air battles, no sieges of heavily fortified positions, no actions from the Korean, Malayan, Algerian, or Vietnamese wars, and has very skimpy coverage of the early World War II (WWII) era battles (1936-1942). The computerized data base has hardly any Asian, African, Mideast, or South American wars (except for a smattering of colonial war battles, and the recent Arab-Israeli wars of 1967 and 1973).

Table 3-1. Scope of the Computerized Data Base

Total number of battles: 601
Battle dates: 1600-1973 A.D.
Total engaged strength: 89×10^6 troops
Total engaged troop-days: 1.1×10^9 troop-days
Total battle casualties: 19×10^6 troops
Average casualty rate: 2 percent per troop-day
Total battle days: 2,300 days
Total distance advanced by ATK: 6,900 km
Total area gained by ATK: 1.3×10^6 sq km

3-3. SUMMARY DISTRIBUTIONS OF SOME KEY VALUES

a. Table 3-2 shows the summary distributions of some key data base values. For example, the attacker's (ATK) recorded strength ranged from 465 to 2,200,000 for the battles in the computerized data base; but half the ATK strengths were less than 23,604, and 5/6 of them were less than 110,000. Also, 1/2 of the recorded ATK strength values were between 13,208 and 70,000, as can be seen from the columns headed 1/4 and 3/4, since 3/4 - 1/4 = 1/2. Similarly, 2/3 of the recorded ATK strength values were between 8,700 and 110,000, as can be seen from the columns headed 1/6 and 5/6. Thus, most battles involved a division to a corps on the attack. Analogous facts can be derived from Table 3-2 for the defender's (DEF) strength, and for the other items listed.

Table 3-2. Summary Distributions of Some Key Data Base Values

	Empirical cumulative distribution for all-HERO data								
Item	MIN	1/6	1/4	1/3	1/2	2/3	3/4	5/6	MAX
ATK strength	465	8,700	13,208	16,225	23,604	47,600	70,000	110,000	2,200,000
DEF strength	155	3,955	6,000	8,141	17,500	35,000	56,000	850,000	1,372,000
ATK casualties	6	258	370	554	1,600	4,000	8,000	15,000	670,000
DEF casualties	1	250	437	730	2,000	4,656	8,036	18,500	885,000
ATK advance (km)	0	0	0.9	1.2	2.6	5.0	7.0	11.6	483
ATK gain (sq km)	0	0	1.5	3.5	13.0	43.5	91.2	264	231,840
Frontal dens (per km)	87	1,500	2,069	2,600	4,000	5,625	6,722	8,529	75,000
Battle date	1600	1793	1820	1863	1915	1942	1943	1944	1973
Duration (days)	1	1	1	1	1	2	3	4	130
Width of front (km)	0.3	2.0	3.0	4.0	7.8	12	18	32	1,060

b. It is interesting to note that half the battles listed occurred after 1915 and half before, and only a few lasted more than 3 or 4 days. Likewise, most battles listed had an attacker's width of front of 2 to 32 km.

c. For completeness, we note that the extreme values (MIN and MAX) in Table 3-2 are associated with the battles listed below. In this list, ISEQNO designates the number of the battle in the computerized data base (see Appendix H for a list of battles in order by ISEQNO). The dates on which these battles began are given in the form YYYYMMDD, i.e., 19421126 means that the year is 1942, the month is 11 (November), and the day is 26 (Thanksgiving Day).

- Attacker strength. Minimum for Chouigi pass, ISEQNO 391, 19421126. Maximum for Vistula-Oder, ISEQNO 511, 194550122.

- Defender strength. Minimum for Medeah Farm, ISEQNO 368, 19181003. Maximum for Defense of Moscow, ISEQNO 489, 19410930.

- Attacker casualties. Minimum for Kilsyth and Majuba Hill (tied), ISEQNO 23 and 232, respectively; 16440815 and 18810227, respectively. Maximum for First Somme, ISEQNO 304, 191660701.

- Defender casualties. Minimum for Tippermuir, ISEQNO 22, 16440901. Maximum for Defense of Moscow, ISEQNO 489, 19410930.

- Attacker advance (km). Minimum for Nieuport, ISEQNO 1, 16000702. (Several other battles were tied with Nieuport for the minimum.) Maximum for Vistula-Oder, ISEQNO 511, 19450112.

- Attacker gain (sq km). Minimum for Nieuport, ISEQNO 1, 16000702. (Several battles were tied with Nieuport for the Minimum.) Maximum for Vistula-Oder, ISEQNO 511, 19450112.

- Frontal density (troops per km). Minimum for Nomonhan Opening Engagement, ISEQNO 259, 1939053. Maximum for Minden, ISEQNO 75, 17590801.

- Battle date. Minimum for Nieuport, ISEQNO 1, 16000702. Maximum for Mount Hermon III, ISEQNO 601, 19731022.

- Duration (days). Minimum for Nieuport, ISEQNO 1, 16000702. (Several other battles were tied with Nieuport for the minimum.) Maximum for Ypres III, ISEQNO 319, 19170731.

- Width of front (km). Minimum for St Amand Farm, ISEQNO 355, 19180718. Maximum for Moscow Counteroffensive, ISEQNO 490, 19411205.

3-4. THE DISTRIBUTION OF BATTLES IN TIME. Figure 3-1 shows the cumulative distribution of battles by date. Note that very few battles in the computerized data base occurred between 1600 and 1620. Then a cluster of battles from the Thirty Years War is listed. Between that period and the era of the War of the Austrian Succession and the Seven Years War only a few battles are listed in the computerized data base, and so forth. Each major war contributed a cluster of battles to the computerized data base. Also, over half of the battles listed before 1900 occurred during either the Napoleonic Wars or the American Civil War.

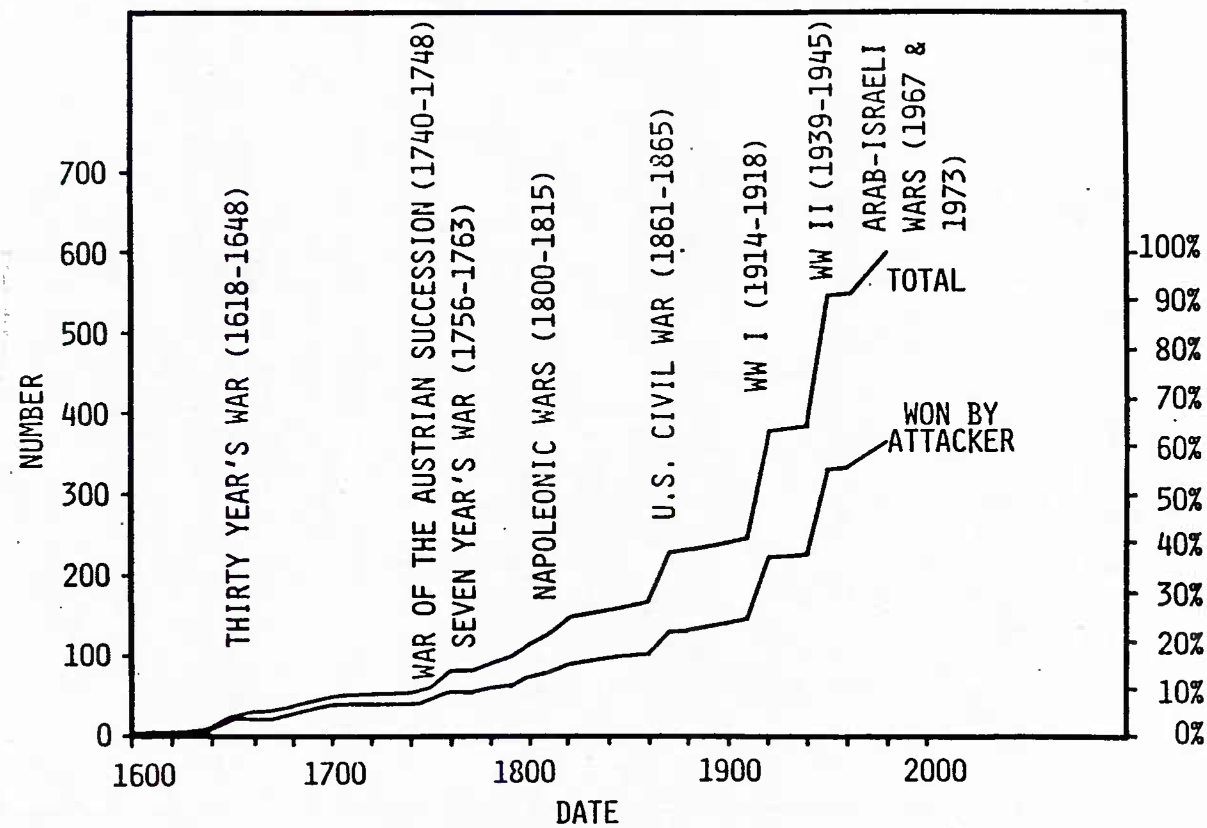


Figure 3-1. Cumulative Distribution of Battles by Date

3-5. FRACTION OF BATTLES WON BY THE ATTACKER

a. Figure 3-2 shows for selected time periods the fraction of battles won by the attacker in the computerized data base. Thus, in the 1600-1699 time period, 36 out of the 48 battles listed in the computerized data base (i.e., 75 percent) were won by the attacker. Superficially, it appears from Figure 3-2 that the fraction of battles won by the attacker decreased gradually from 1600 to just before 1900, and thereafter rose somewhat; and perhaps it did. But the statistical confidence bands on the average fractions are so broad that the data are also consistent with the assumption that the fraction of the battles won by the attacker has remained constant at about 61 percent over the entire time period 1600-1979. This is shown in Figure 3-2 by the fact that all of the confidence bands overlap, usually by fairly wide margins, the line at 61 percent.

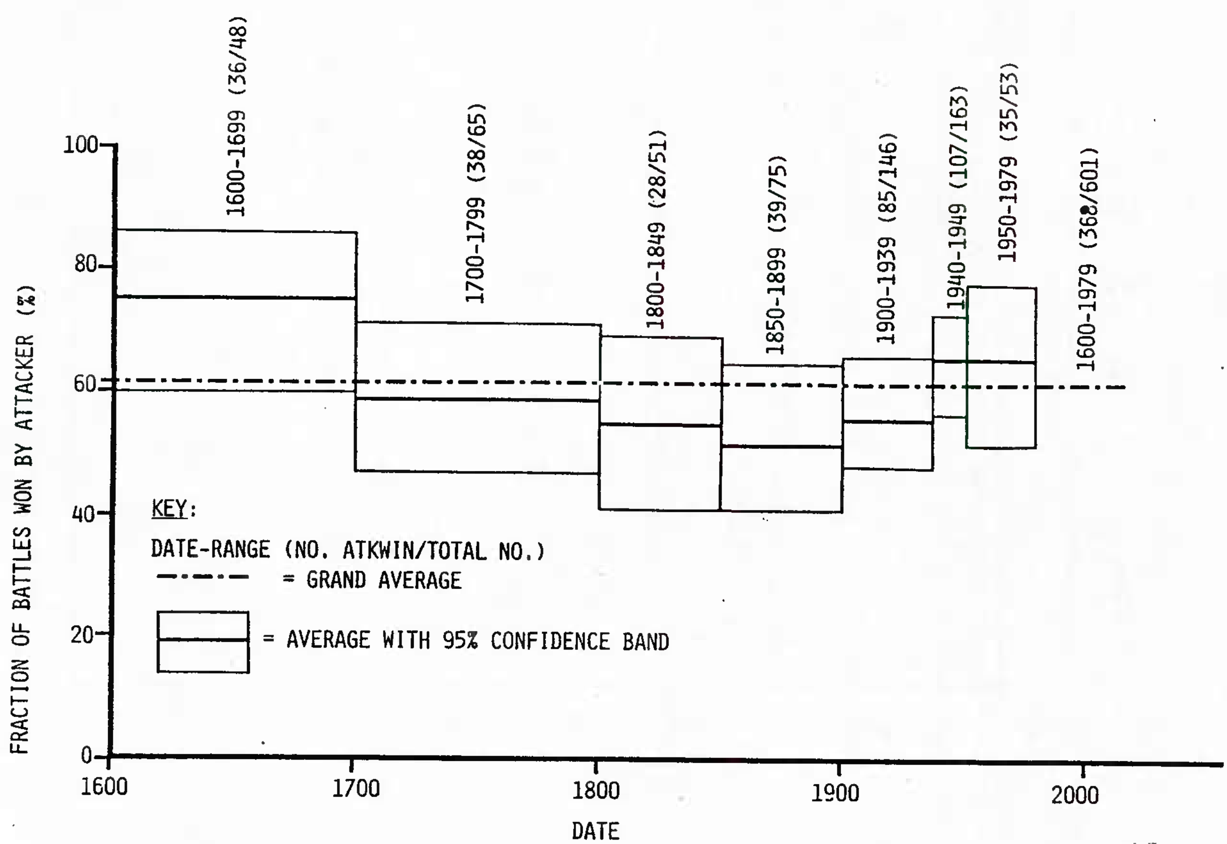


Figure 3-2. Fraction of Battles Won by Attacker versus Time Period

b. Table 3-3, showing battle outcome versus time period, was prepared to examine this issue in more detail. The chi-square test for independence in contingency tables **(Ref 3-1 and 3-2)**, applied to Table 3-3, indicates that the significance level is a little over 10 percent. So, the evidence in favor of a secular change in the probability of an attacker victory is too slight to be depended upon.

Table 3-3. Battle Outcome versus Time Period

Time period	Number of battles (percent of row total)[a]		
	ATKWIN	DRAW or DEFWIN	Total
1600-1699	36 (75.00)	12 (25.00)	48 (100.00)
1700-1799	38 (58.46)	27 (41.54)	65 (100.00)
1800-1849	28 (54.90)	23 (45.10)	51 (100.00)
1850-1899	39 (52.00)	36 (48.00)	75 (100.00)
1900-1939	85 (58.22)	61 (41.78)	146 (100.00)
1940-1949	107 (65.64)	56 (34.36)	163 (100.00)
1950-1979	35 (66.04)	18 (33.96)	53 (100.00)
Total	368 (61.23)	233 (38.77)	601 (100.00)

[a]Percentages may not sum to total due to rounding. Chi-square = 10.01 at 6 degrees of freedom, which is significant at a little over the 12 percent level.

3-6. DISTRIBUTION OF BATTLE DURATIONS

a. As noted in Appendix I, the HERO data base gives battle durations (T) in units of days, which is too coarse a time scale to be useful for many purposes, including any sophisticated statistical work on battle durations. However, it is of interest to obtain a descriptive distribution of battle durations for the computerized data base. This was done by a trial-and-error process of fitting alternative distributions to the empirical distribution of battle durations.

b. Table 3-4 shows that the battle durations can be rather closely fitted by Weibull distributions with an offset of 1/2 day. This offset may be caused by the coarseness of the time scale. This is an intriguing finding since the Weibull distribution is often used as a distribution of time to failure in reliability engineering. Weibull distributions have also been reported to fit the distribution of the durations of battle and nonbattle personnel disablement periods (see the Editor's Introduction to Reference 3-3), industrial strikes (Ref 3-4), and wars (see Ref 3-4 and the Editor's Introduction to Ref 3-3).

Table 3-4. Comparison of Battle Duration Distributions

Duration T (days)	Percent of all battles with durations not exceeding T							
	All-HERO battles				Non-WW II battles			
	Empirical	Fitted* Weibull	Fitted** lognormal	Fitted*** exponential	Empirical	Fitted+ Weibull	Fitted++ lognormal	Fitted+++ exponential
1	52.1	53.7	52.2	34.0	66.8	63.2	66.0	35.5
2	68.4	70.5	68.4	56.5	77.8	76.4	77.1	58.3
3	78.5	78.0	76.7	71.3	82.6	81.9	82.6	73.1
4	83.9	82.5	81.7	81.1	86.5	85.2	85.9	82.6
5	87.5	85.6	85.2	87.5	88.8	87.5	88.2	88.8
6	89.7	87.8	87.6	91.8	89.9	89.2	89.9	92.8
7	90.7	89.6	89.5	94.6	90.8	90.5	91.2	95.3
8	92.3	91.0	90.9	96.4	92.2	91.5	92.2	97.0
9	92.5	92.1	92.1	97.6	92.4	92.4	93.0	98.1
10	92.8	93.0	93.0	98.4	92.7	93.1	93.7	98.7
12	94.0	94.4	94.5	99.3	93.8	94.2	94.7	99.5
14	94.8	95.4	95.5	99.7	94.7	95.0	95.5	99.8
16	95.8	96.2	96.2	99.9	96.1	95.7	96.1	99.9
18	96.3	96.8	96.8	99.9	96.6	96.2	96.6	100.0
20	96.8	97.2	97.3	100.0	96.8	96.6	96.9	100.0
25	97.8	98.1	98.1	100.0	97.7	97.4	97.6	100.0
50	99.5	99.5	99.4	100.0	99.6	99.0	99.0	100.0
100	99.7	99.9	99.9	100.0	99.7	99.7	99.7	100.0

$$*F(T) = 1 - EXP\left[-\left(\frac{T-1/2}{0.93}\right)^{0.42}\right]$$

$$+F(T) = 1 - EXP\left[-\left(\frac{T-1/2}{1/2}\right)^{1/3}\right]$$

$$**F(T) = CUNO\left(\frac{LOG(T/0.91)}{1.646}\right)$$

$$++F(T) = CUNO\left(\frac{LOG(T/0.42)}{2.11}\right)$$

$$***F(T) = 1 - EXP(-0.416T)$$

$$+++F(T) = 1 - EXP(-0.438T)$$

c. Although the Weibull distribution has a strong theoretical appeal because of its connection with the theory of reliability, it is nevertheless true that a lognormal distribution adequately fits the battle duration data, as shown in Table 3-4, Figure 3-3, and Figure 3-4. The lognormal distribution also fits the duration data on wars and industrial strikes about as well as the Weibull distribution does **(Ref 3-5)**. However, as Figures 3-3 and 3-4 show, an exponential distribution is a much worse fit to the data than either the Weibull or the lognormal distribution.

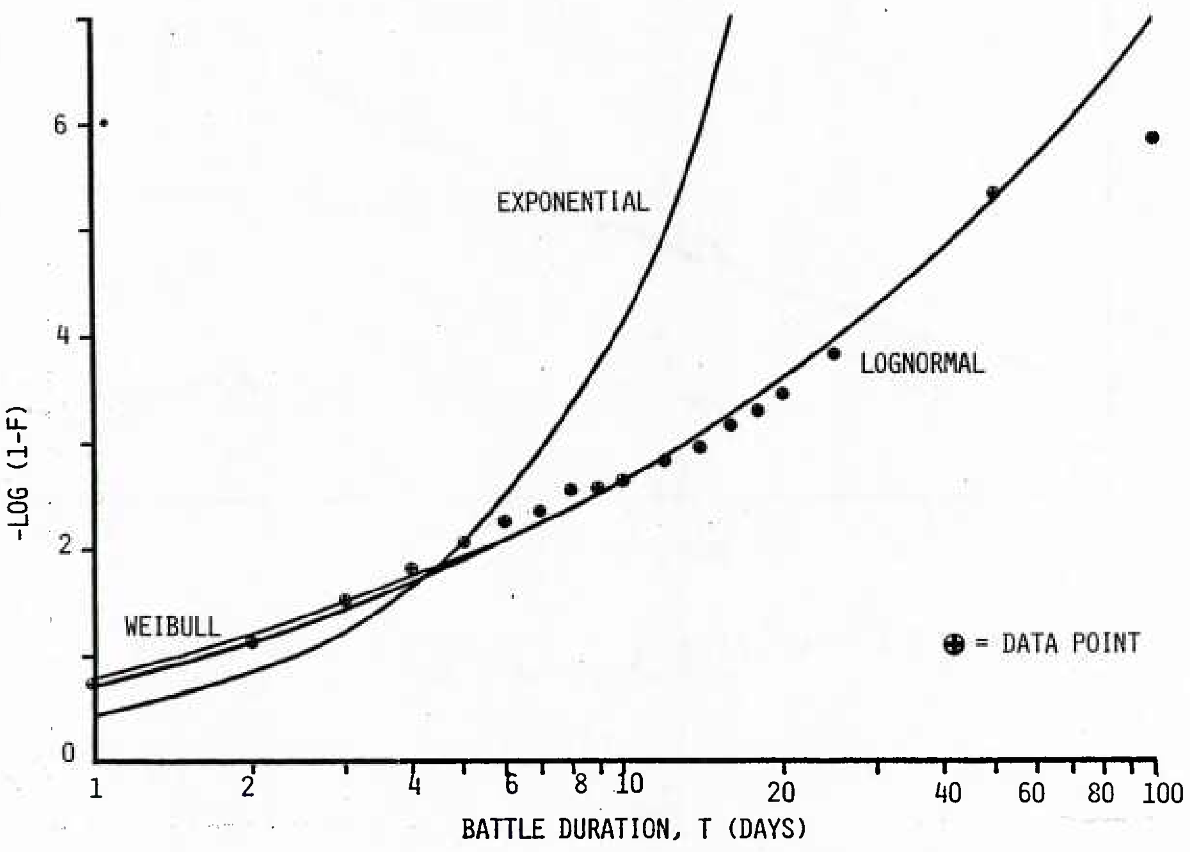


Figure 3-3. Comparison of Battle Duration Distributions for All-HERO Data

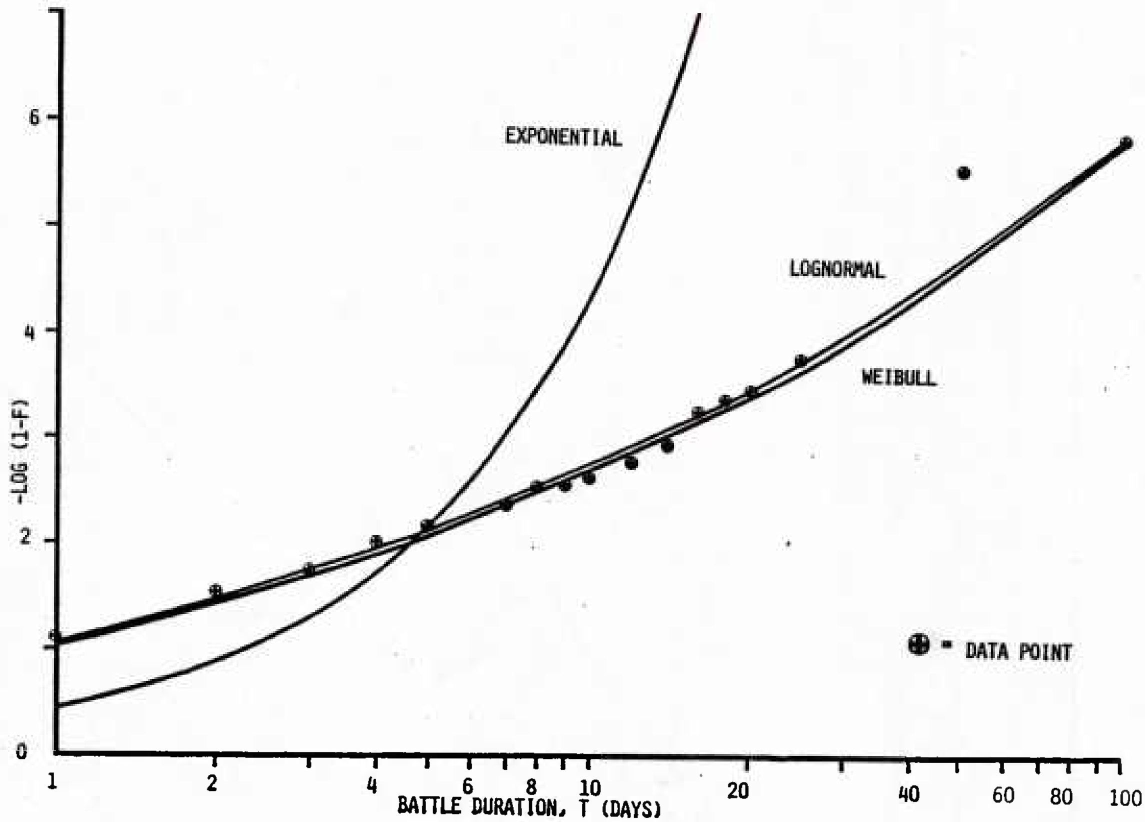


Figure 3-4. Comparison of Battle Duration Distributions for Non-WWII Subsample

d. For the durations of wars, Weiss **(Ref 3-6)** has derived a distribution entirely different from those cited above. No attempt was made in this phase of the CHASE Study to fit Weiss's form of distribution to the battle duration data.

3-7. DISTRIBUTIONS OF SOME OTHER SELECTED QUANTITIES

a. Description of Quantities Selected

(1) We will present the empirical distributions of the following quantities:

- Attacker's personnel casualty fraction, FX
- Defender's personnel casualty fraction, FY
- Attacker's personnel force ratio, FR
- Defender's personnel casualty exchange ratio, CER
- Attacker's adjudged mission accomplishment rating, ACHA
- Defender's adjudged mission accomplishment rating, ACHD

(2) Except for ACHA and ACHD, these quantities are not given directly by the information in the data base, but are derived from directly-given quantities. Their definitions are as follows (see also the Glossary):

- FX = CX / XO
- FY = CY / YO
- FR = XO / YO
- CER = CX / CY
- FER = FX / FY

Clearly, FER may also be written in the mathematically equivalent form FER = CER / FR.

b. Summary Distributions. Table 3-5 gives the summary distributions of FX, FY, FR, CER, and FER. We observe from this tabulation that in the computerized data base battles, the defender's casualty fraction tends to be larger than the attacker's. For example, the median values of FX and FY are about 7.1 and 12.3, respectively, and FY is roughly double the FX at the same cumulative probability level. We also observe that casualty fractions in excess of 20 or 30 percent do occur in these battles, but that they are rare. The median FR value is about 1.5. Also, we see from Table 3-5 that the attacker was outnumbered (that is, FR less than 1.0) in about 1/3 of the battles in the computerized data base, was in fact outnumbered by better than 5 to 4 (that is, FR less than 0.8) in over 1/6 of those battles, and was able to achieve better than a 3 to 1 force ratio (FR greater than 3.0) in about 1/6 of the battles. In about 2/3 of the battles, the force ratio was between 0.8 and 3.0. As shown by Table 3-5, the median CER is about 1.0, indicating that the attacker's casualties outnumber the defender's (that is, CER greater than 1.0) in about half of these battles. As shown by Table 3-5, the attacker's personnel casualties are 1/2 to 2 times the defender's in about half the battles. They are between 1/3 and 3 times the defender's casualties in about 2/3 of the battles. Although either the attacker's or the defender's casualty exchange ratio reportedly exceeds 100 to 1 for some battles, these values strain one's credulity. Note that the personnel losses CX and CY are not supposed to include prisoners taken in pursuit after the main battle has ended (see Appendix E, paragraph E-2c(2)). Some of the FER values also seem incredibly high or low.

Table 3-5. Summary Distributions of Some Selected Quantities

Quantity	Empirical cumulative distribution for All-HERO data								
	MIN	1/6	1/4	1/3	1/2	2/3	3/4	5/6	MAX
FX (percent)	0.122	1.750	2.546	3.529	7.065	12.141	16.129	21.818	84.455
FY (percent)	0.033	3.512	5.085	7.292	12.282	20.688	28.266	36.296	100.00
FR	0.236	0.809	0.948	1.077	1.522	2.031	2.438	3.034	20.1530
CER	0.001	0.267	0.400	0.567	0.966	1.500	1.953	2.752	3,000.0
FER	0.00139	0.158	0.242	0.336	0.619	1.050	1.357	1.889	1,323.5

c. **Graphical Distributions**

(1) **Distributions Other than Achievement Scores.** Figures 3-5 through 3-8 provide graphical distributions for FX, FY, FR, CER, and FER. Note that these empirical distribution functions are plotted on lognormal probability scales. This was done to improve the linearity of the plotted empirical distribution curves. The straight lines shown in Figures 3-6 through 3-8 were fitted by eye to the empirical distribution functions. These graphs suggest that FX, FY, FR, CER, and FER are approximately lognormally distributed (see **Ref 3-7** for a description of the lognormal distribution).

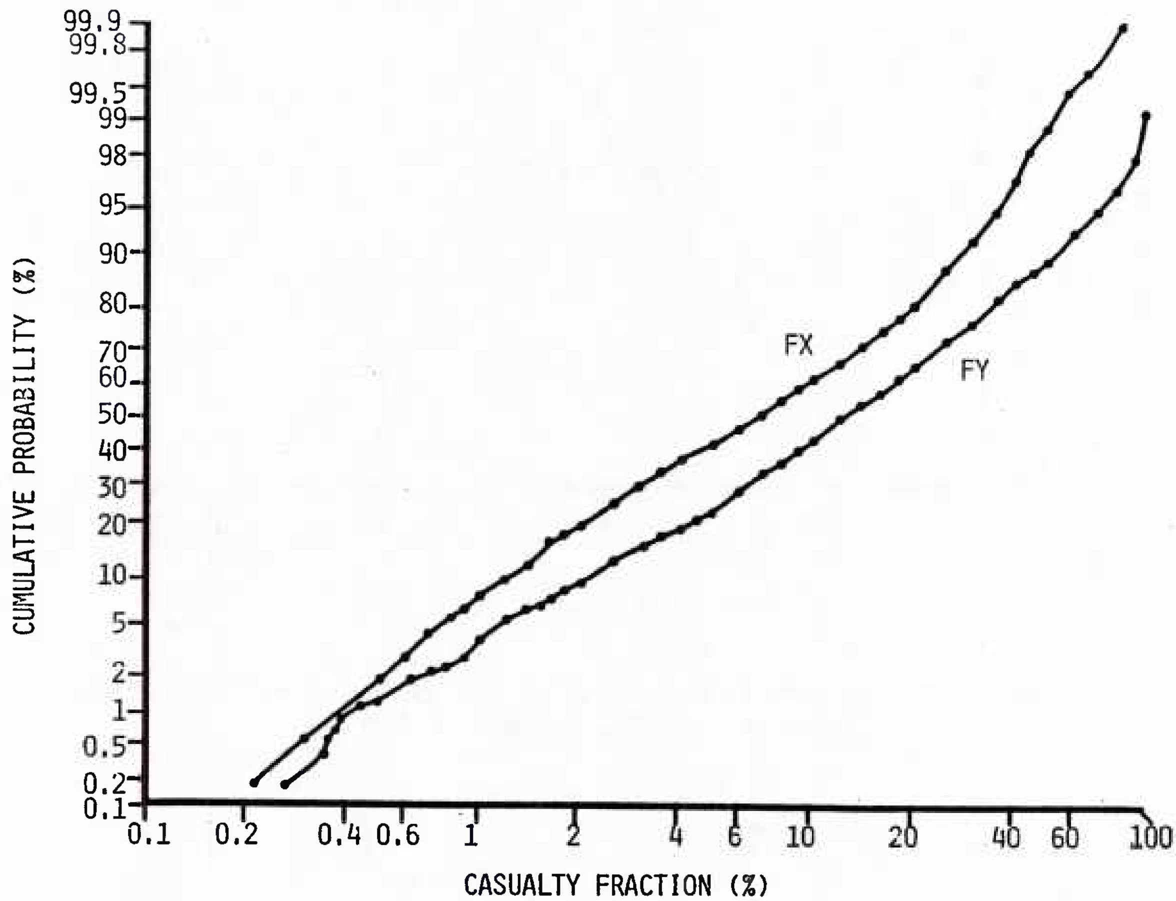


Figure 3-5. Empirical Distribution of Personnel Casualty Fractions for the Attacker (FX) and for the Defender (FY), Using All-HERO Data

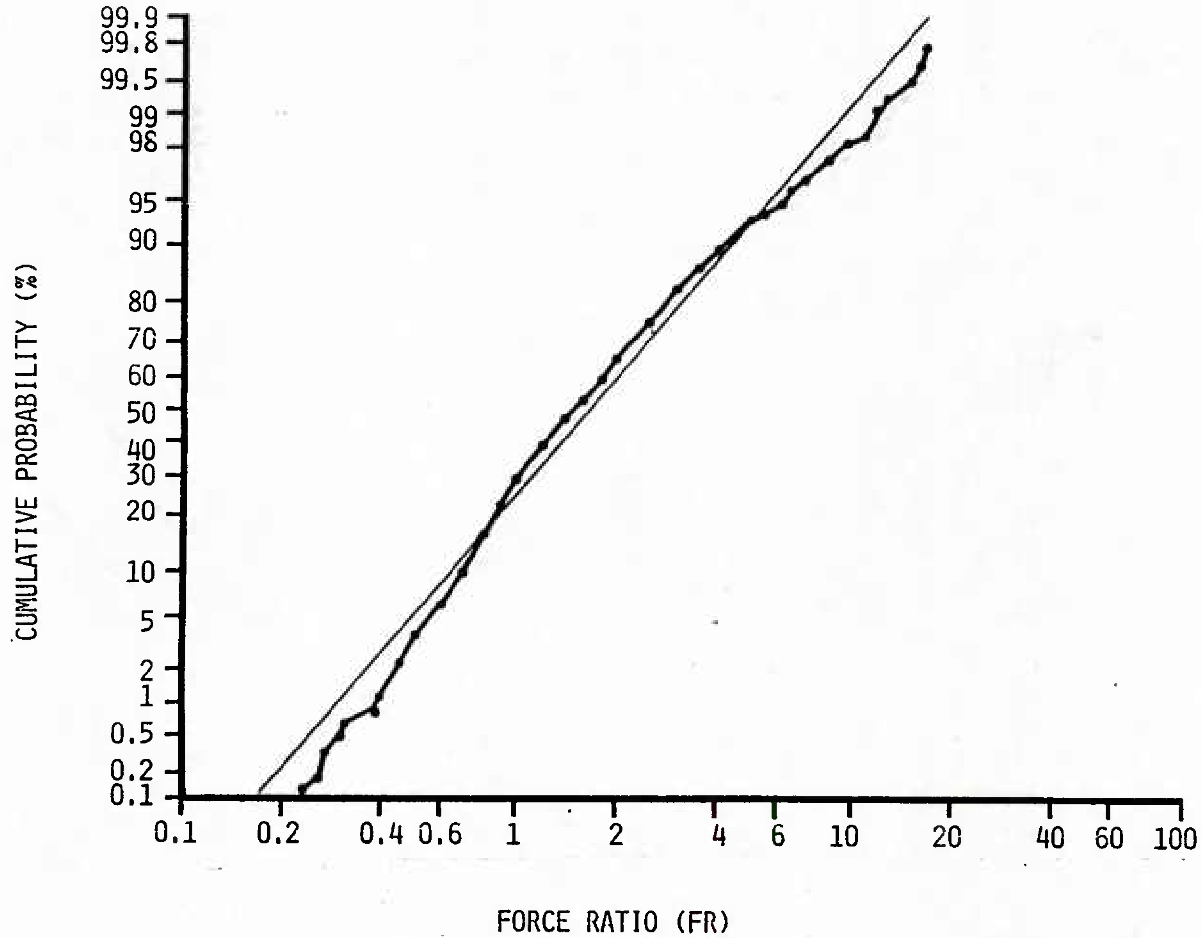


Figure 3-6. Empirical Distribution of the Attacker's Personnel Force Ratio (FR), Using All-HERO Data

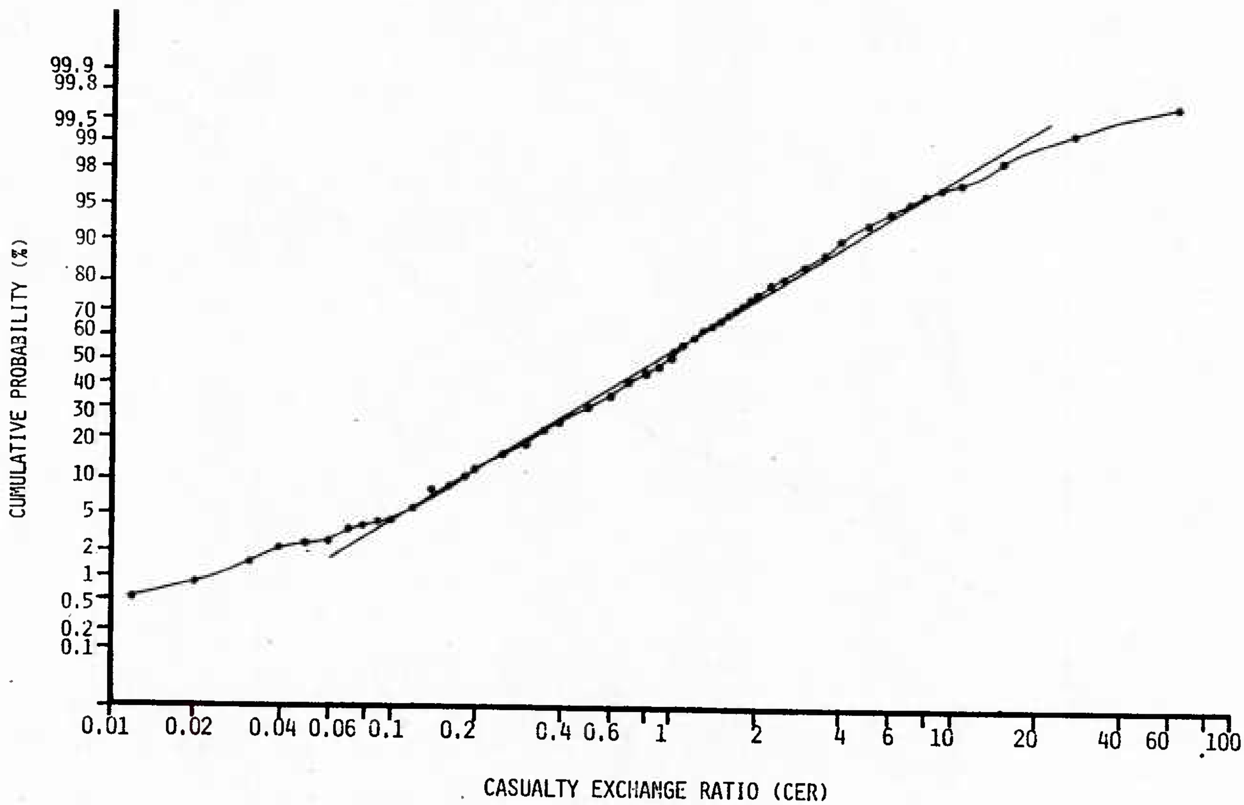


Figure 3-7. Empirical Distribution of the Defender's Personnel Casualty Exchange Ratio (CER), Using All-HERO Data

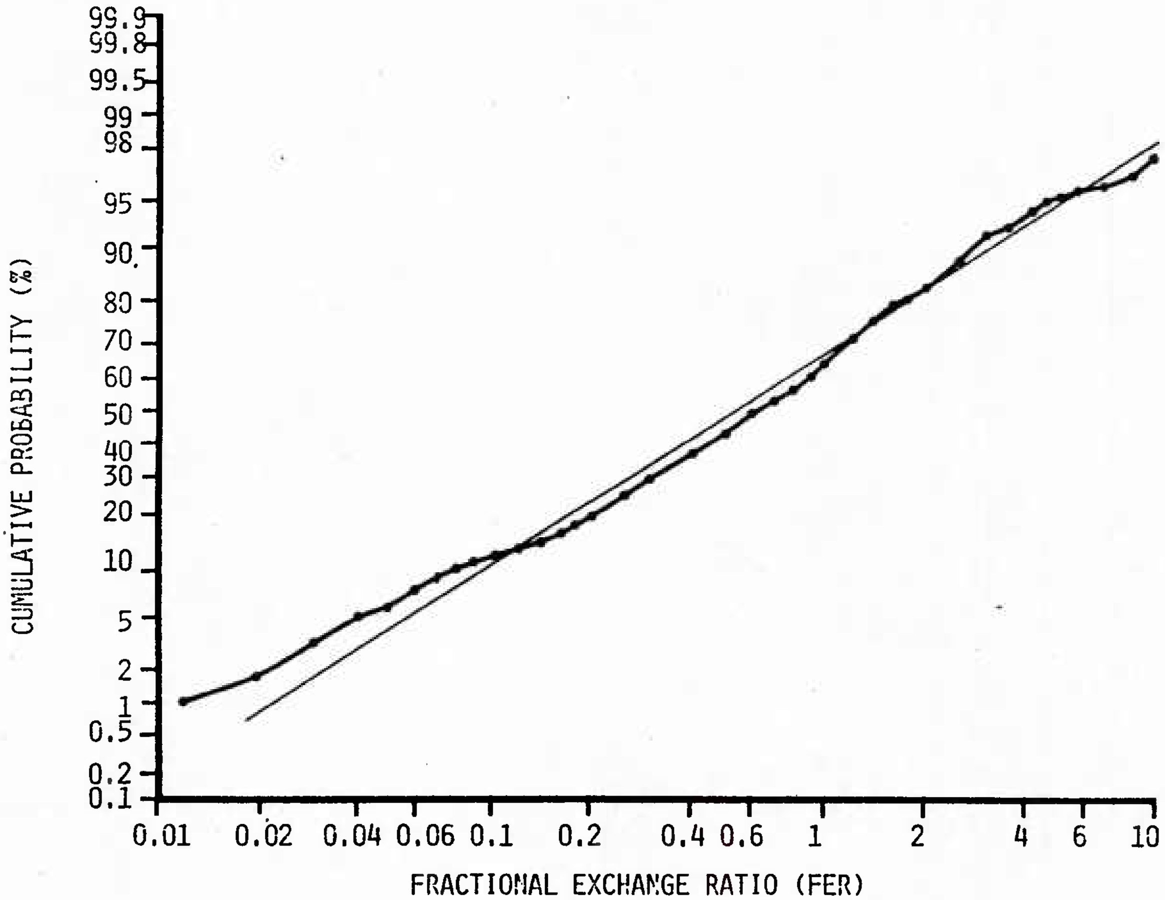


Figure 3-8. Empirical Distribution of the Defender's Personnel Fractional Exchange Ratio (FER), Using All-HERO Data

(2) Since FX, FY, FR, CER, and FER may be approximately lognormally distributed, it is appropriate to present some descriptive statistics for the distributions of their logarithms. This is done in Table 3-6. The headings in this table have the following significance:

- MEAN is the average value of the quantity for the battles in the computerized data base (Ref 3-8).
- S.D. is the standard deviation of the quantity for the battles in the computerized data base (Ref 3-8).
- SKEW is the coefficient of skewness (see Glossary and Refs 3-1, 3-8).
- XKURT is the coefficient of excess kurtosis, sometimes called simply the excess (see Glossary and Refs 3-1, 3-8).
- MIN and MAX are the minimum and maximum values of the quantity in the computerized data base. The table gives the MIN and MAX values, and the ISEQNOs of the battles at which the MIN and MAX values occur. See Appendix H for an index of battles by ISEQNO.
- Sample Size is the number of battles on which the MEAN and S.D. values are based (Ref 3-2).
- PROB. KOLMOG. EXCEEDANCE is the probability that the Kolmogoroff test criterion is exceeded (see Glossary and Refs 3-2, 3-8, 3-9). The Kolmogoroff test is also sometimes called the Kolmogoroff-Smirnov test.

Table 3-6. Descriptive Statistics of Some Selected Quantities Using All-Hero Data[a]

Quantity	MEAN	S.D.	SKEW	XKURT	MIN		MAX		Sample size	PROB. KOLMOG. EXCEEDANCE (percent)
					Value	ISEQNO	Value	ISEQNO		
LOG(FX)	-2.777	1.201	-0.322	-0.562	-6.705	23	-0.169	92	583	2.9
LOG(FY)	-2.178	1.238	-0.589	0.467	-8.006	22	0.000	78	583	16.9
LOG(FR)	0.466	0.728	0.544	0.432	-1.372	531	3.003	371	598	6.0
LOG(CER)	-0.132	1.361	-0.057	3.675	-6.908	23	8.006	22	583	15.1
LOG(FER)	-0.599	1.482	-0.190	2.019	-6.580	23	7.188	22	583	7.7

[a]See text, paragraph 3-7c(2), Chapter 3, for an explanation of the column headings.

The PROB. KOLMOG. EXCEEDANCE values provide a measure of how close the empirical distribution function is to being lognormal. Specifically, they indicate that the empirical distributions of FY and CER are approximately lognormal, that the empirical distributions of FR and FER may be only marginally lognormal, and that the empirical distribution of FX is statistically significantly different from lognormal.

(3) Distributions of the Attacker's and Defender's Achievement Scores. Figure 3-9 presents the distributions of the attacker's and defender's achievement ratings. These quantities are symbolized by ACHA and ACHD, respectively. As explained in Appendices E and F, they are ratings on a scale of 0 (unsuccessful) to 10 (fully successful) of the extent to which the respective sides were able to accomplish their missions. From Figure 3-9, it is evident that, on the average, the attacker is rated higher in mission accomplishment than the defender--which is consistent with scoring the attacker as the victor in 61 percent of the battles, as mentioned in paragraph 3-5. As shown by the relative lengths of the bars in Figure 3-9, the attacker is credited much more frequently than the defender with an achievement rating of 8, 9, or 10. Similarly, the defender is given much more frequently than the attacker an achievement rating of 2, 3, or 4.

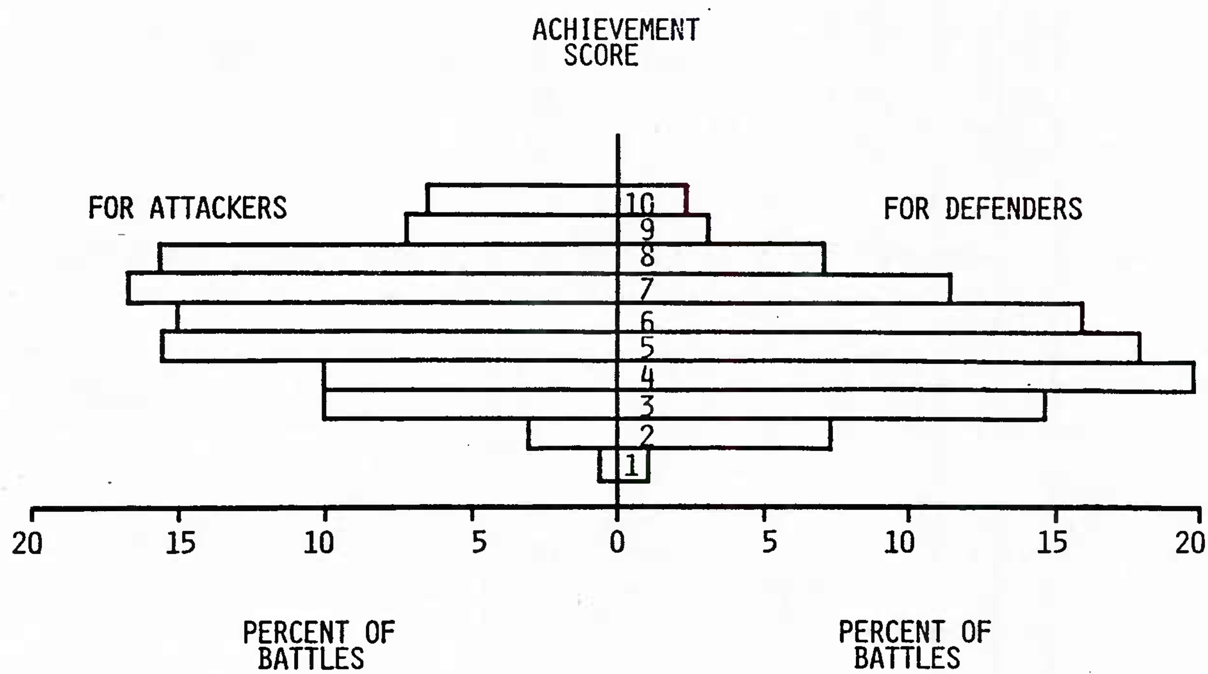


Figure 3-9. Histogram of Achievement Scores for Attackers and Defenders

3-8. THE DEPENDENCE OF VICTORY ON FORCE RATIO

a. The question of the extent to which victory in battle is dependent on force ratio has been contemplated by many students of military history and science. Many of them have argued that force ratio has a strong, almost conclusive influence on the outcome. This view is represented by such aphorisms as "Get thar fustest with the mostest," "God is always on the side of the big battalions," "Place the maximum force at the decisive point," and so forth. Clausewitz **(Ref 3-10)**, in Book 3, Chapter 8, immediately after citing the examples of Leuthen, Rossbach, Dresden, Kolin, and Leipzig--all of which were fought either by Frederick the Great or by Napoleon--states flatly that, "These examples may show that in modern Europe even the most talented general will find it very difficult to defeat an opponent twice his strength. When we observe that the skill of the greatest commanders may be counterbalanced by a 2 to 1 ratio in the fighting forces, we cannot doubt that in ordinary cases, whether the engagement be great or small, a significant superiority in numbers (it does not have to be more than double) will suffice to assure victory however adverse the other circumstances. ... The first rule, therefore, should be: put the largest possible army into the field." In a similar vein, General Depuy states **(Ref 3-11)** that, "Conventional military wisdom has long had it that a defender can cope with a 3 to 1 adverse force ratio. ... Conventional wisdom, based on experience, is supported by wargaming and analysis. Over a long period, the wargames conducted at Ft. Leavenworth, Kansas, the Combined Arms Center of the US Army, affirm that the defender usually begins to lose when the attacker's advantages rise above 3 to 1. ... At the Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland, the threshold is 2.6 to 1. So, 3 to 1 is a good round figure." Nevertheless, several analyses applying quantitative methods to historical combat data found only a weak dependence of victory on force ratio **(Refs 3-12, 3-13, 3-14, 3-15, 3-16, and 3-17)**.

b. To determine what light might be shed on this issue by the computerized data base, we constructed Table 3-7, displaying battle outcomes versus various ranges of force ratio. The chi-square test for independence in contingency tables (Refs 3-1, 3-2) applied to Table 3-7, indicates that the significance level is about 4 percent. Hence, battle outcomes do indeed depend on force ratio.

Table 3-7. Battle Outcome versus Force Ratio for All-HERO Data

Force ratio	Number of battles (percent of row total)[a]			
Range	DEFWIN	DRAW	ATKWIN	Total
Less than 1/3	2 (40.0)	0 (0.0)	3 (60.0)	5 (100.00)
Between 1/3 and 2/3	23 (51.1)	1 (2.2)	21 (46.7)	45 (100.00)
Between 2/3 and 3/2	88 (35.8)	16 (6.5)	142 (57.7)	246 (100.00)
Between 3/2 and 3	60 (30.6)	12 (6.1)	124 (63.3)	196 (100.00)
Greater than 3	23 (21.7)	5 (4.7)	78 (73.6)	106 (100.00)
Total	196 (32.8)	34 (5.7)	368 (61.5)	598 (100.00)

[a]Percentages may not sum to total due to rounding. Chi-square = 16.18 at 8 degrees of freedom, which is significant at about the 4.0 percent level.

c. However, the degree of dependence is by no means as marked as some might have expected. For example, although the attacker wins about 74 percent of the battles in which the force ratio is at least 3, he also wins about 62 percent of the battles regardless of whether the force ratio is favorable or not. Hence, a force ratio of 3 raises the attacker's chance of winning from about 62 percent to about 74 percent. No doubt this is a worthwhile increase, and one the attacker is surely loath to forego, but it is far from assuring a victory by the attacker. Nor is it by any means necessary for the attacker to muster a 3 to 1, or even a 2 to 1, advantage to win. Table 3-7 shows that the attacker's chance of winning is still close to 50 percent even for FR values between 1/3 and 2/3, that is, when the attacker is outnumbered by between 1 to 3 and 2 to 3.

d. That there is a statistically significant, but only a weak and not particularly reliable dependence of battle outcome on force ratio, is a finding that supports and confirms the earlier quantitative analyses cited in the preceding paragraph. The search for factors associated with victory is continued in Chapter 4.

3-9. NEXT STEPS FOR DESCRIPTIVE STATISTICS. The present findings are but a token of the descriptive statistics that could be developed. Table 3-8 lists some of the desirable next steps for descriptive statistics work.

Table 3-8. Next Steps for Descriptive Statistics

- Recalculate and revise the descriptive statistics as the CDES results become available
- What data to trust, include, or treat separately hinges on resolution of the WWII anomaly (see Chapter 4, paragraph 4-4 for a description of the WWII anomaly)
- Add distributions of rates (of advance, of losses, etc.) as CDES provides more precise data on battle durations
- Plot selected values versus battle date
- Correlate and cross-plot pairs of variables, e.g.,
  - -- The two measures of surprise (SURPA and SURPAA)
  - -- Maneuver (MANA) and linear troop density
  - -- Casualties (CX and CY)
- Look for connections between the subjective and objective assessments, e.g., subjective terrain favoring attacker (TERRA) vs
  - -- Objective terrain descriptors (TERRA1/TERRA2)
  - -- Objective weather descriptors (WX1/WX2/WX3)
- Try to fit functions to various distributions, e.g.,
  - -- Are the attacker and defender casualty fractions (FX and FY) Weibull-distributed?
  - -- Is the force ratio (FR) lognormally distributed?
  - -- Is battle duration (T) distributed according to Weiss's formula?
- Look for interrelationships among variables, e.g., between losses and battle duration
- What can be said about losses of heavy equipment (such as armor, artillery, air)
- Interpret and document findings

3-10. CONCLUDING OBSERVATIONS ON DESCRIPTIVE STATISTICS

a. Descriptive statistics express succinctly the predominant characteristics of a mass of data and provide insights that usefully supplement those obtained by a study of individual cases. However, a clear perception of cause and effect relationships usually requires more sophisticated techniques.

b. The HERO data base is mainly representative of short, pitched land combat battles fought by organized division- and corps-sized military formations during the 19th and early 20th centuries in Europe and North America.

c. The attacker won about 61 percent of the 601 battles recorded in the HERO data base. The probability of an attacker victory may have declined slightly from 1600 to about 1850-1900, and then risen between 1850-1900 to the 1970s, but the evidence for this gradual secular change is too slight to be depended upon.

d. Battle durations seem to be distributed approximately as Weibull or lognormal random variables.

e. The defender's personnel casualty fraction tends to be larger than the attacker's.

f. The attacker's personnel force ratio seems to be distributed roughly as a lognormal random variable. The attacker outnumbers the defender by a 3 to 1 margin in only about one-sixth of the battles. Victory seems to depend somewhat on force ratio, but not in a particularly reliable way. A 3 to 1 force ratio is neither necessary nor sufficient to assure a victory in a battle.

g. The defender's personnel casualty exchange ratio is distributed approximately as a lognormal random variable. Since its median value is close to unity, the attacker's personnel casualties outnumber the defender's in about half the battles.

h. The defender's personnel fractional exchange ratio seems to be distributed roughly as a lognormal random variable. It is less than unity in about two-thirds of the battles.

CHAPTER 4

FACTORS ASSOCIATED WITH VICTORY

4-1. INTRODUCTION

a. Scope and Objectives. This chapter presents an initial analysis of the factors associated with victory. It can be considered as an early stage in the refinement and expansion of the discussion of the dependence of victory on force ratio in Chapter 3, paragraph 3-8. This work is motivated partly by the desire to uncover all of the important causes of victory in battle. However, it is also motivated by the following important technical statistical considerations. Many of the statistical techniques intended for subsequent use in CHASE require variables that are one-dimensional, continuous, unbounded above and below, and equipped with a measure of the distance between two different values. Yet the conventional designation of battle outcomes as wins and losses (or as wins, losses, and draws) provides only a discontinuous and bounded variable that, while one-dimensional, is not equipped with any evident measure of the distance between two different values. Thus, a main goal of this preliminary analysis is to find at least one variable that is:

(1) One-dimensional.

(2) Continuous.

(3) Unbounded above and below.

(4) Equipped with a measure of the distance between two different values.

(5) Sufficiently representative of the conventional win, lose, or draw categories of battle outcome that it can be substituted for them in later statistical analyses.

b. Outline of Approach. Each of the following six variables will be considered for suitability as a surrogate for the conventional battle outcome categories:

(1) Force ratio (FR)

(2) Bitterness (EPS)

(3) Casualty exchange ratio (CER)

(4) Fractional exchange ratio (FER)

(5) Advantage (ADV)

(6) Residual advantage (RESADV)

Each of these variables can be defined objectively and quantitatively in terms of the initial personnel strengths and losses to the engaged sides, as shown in paragraph 4-2, below. Thus, all of them are determined by objective numerical data rather than by subjective or qualitative data. In addition, each of them (possibly after taking their logarithms, as in the case of FR, EPS, CER, and FER) satisfies criteria (1) through (4), above. Thus, (5) is the only criterion that remains to be addressed. In this paper, logistic regression is the principal technique used to assess the degree to which the surrogate variables are representative of the conventional battle outcome categories (win, lose, or draw). Logistic regression--not to be confused with logarithmic regression--is a statistical method that is widely used for similar purposes in traffic flow, safety, toxicology, pharmacology, economics, sociology, and other disciplines. Appendix J provides an introduction to the theory of this technique. For additional related material see **Refs 4-1, 4-2, and 4-3.** However, before applying logistic regression, we need to define some of the candidate variables (particularly ADV, EPS, and RESADV) and to indicate why they are included as possible surrogates for the conventional battle outcome categories.

4-2. DEFINITION AND EMPIRICAL DETERMINATION OF CANDIDATE VARIABLES

a. Orientation. The variables FR, CER, and FER were defined in Chapter 3, paragraph 3-7, and do not require further explanation. The variables ADV and EPS arise naturally from a consideration of Lanchester's square-law equations, and RESADV is defined in terms of ADV and FR. Accordingly, we begin with a consideration of Lanchester's equations which we write in the form:

$$dX/dt = - DD * Y \qquad (4\text{-}1.1)$$

$$dY/dt = - AA * X \qquad (4\text{-}1.2)$$

$$X(0) = X0 \qquad (4\text{-}1.3)$$

$$Y(0) = Y0 \qquad (4\text{-}1.4)$$

where X = X(t) and Y = Y(t) are the attacker's and the defender's surviving personnel strengths at time t into the battle, X0 and Y0 are the attacker's and defender's initial personnel strengths, and AA and DD are the attacker's and the defender's personnel activity parameters that measure the rate at which they inflict losses on the opposing side (in number of opponents lost per friendly troop per unit time). The following discussion of these equations is based on material in **Refs 4-4, 4-5, and 4-6.**

b. Solution of Lanchester's Equations. A general scientific principle is to consolidate two or more variables into one dimensionless quantity in order to simplify the problem by reducing the number of variables that need to be addressed. To apply this principle to Lanchester's equations, divide the strengths by their initial values to write Equations (4-1) as:

$$dA/dt = - DELTA * D \tag{4-2.1}$$

$$dD/dt = - ALPHA * A \tag{4-2.2}$$

$$A(0) = 1 \tag{4-2.3}$$

$$D(0) = 1 \tag{4-2.4}$$

where:

$$A = X / XO \tag{4-3.1}$$

$$D = Y / YO \tag{4-3.2}$$

$$ALPHA = AA * XO / YO \tag{4-3.3}$$

$$DELTA = DD * YO / XO \tag{4-3.4}$$

The solution of Equations (4-2) can be written as:

$$A = COSH(EPS) - MU * SINH(EPS) \tag{4-4.1}$$

$$D = COSH(EPS) - MU^{-1} * SINH(EPS) \tag{4-4.2}$$

where:

$$MU = SQR\ (DELTA / ALPHA)$$

$$= (YO / XO) * SQR\ (DD / AA) \tag{4-5}$$

$$EPS = T * LAMBDA \tag{4-6}$$

$$LAMBDA = SQR\ (ALPHA * DELTA) = SQR\ (AA * DD) \tag{4-7}$$

and:

T = Duration of the battle, in time units.

c. Theoretical Interpretation of the Parameters Appearing in the Solution of Lanchester's Equations. The parameters in question are EPS, MU, and LAMBDA, where EPS and LAMBDA are related as in Equation (4-6). As will now be explained, these parameters are of important theoretical significance. Moreover, as is shown in paragraph 4-2d, their values for a battle can be estimated from historical data. The empirical values of these parameters will play a large role throughout the remainder of this chapter. By Equation (4-7), LAMBDA is the geometric mean of the attacker's and the defender's activity coefficients in a battle and has the dimensions of a rate. Accordingly, LAMBDA is an index of the average rate at which the casualty fractions increase during a battle, and so will be called the intensity of the battle. Then EPS, being by Equation (4-6) the product of an average rate by the time over which it persists, is an index of the total casualty fraction incurred over the whole course of the battle. Hence, it will be called the bitterness of the battle (see also Equation (4-12.2)). The value of MU determines which side has the upper hand, in the sense that:

(1) If MU is greater than 1, then A theoretically goes to zero before D does and so the defender has the upper hand.

(2) If MU is less than 1, then D theoretically goes to zero before A does and so the attacker has the upper hand.

Accordingly, we define the (defender's) advantage to be:

$$ADV = LOG(MU), \tag{4-8}$$

so that the defender theoretically has the advantage when ADV is greater than zero but is at a disadvantage relative to the attacker when ADV is less than zero, as illustrated in Figure 4-1.

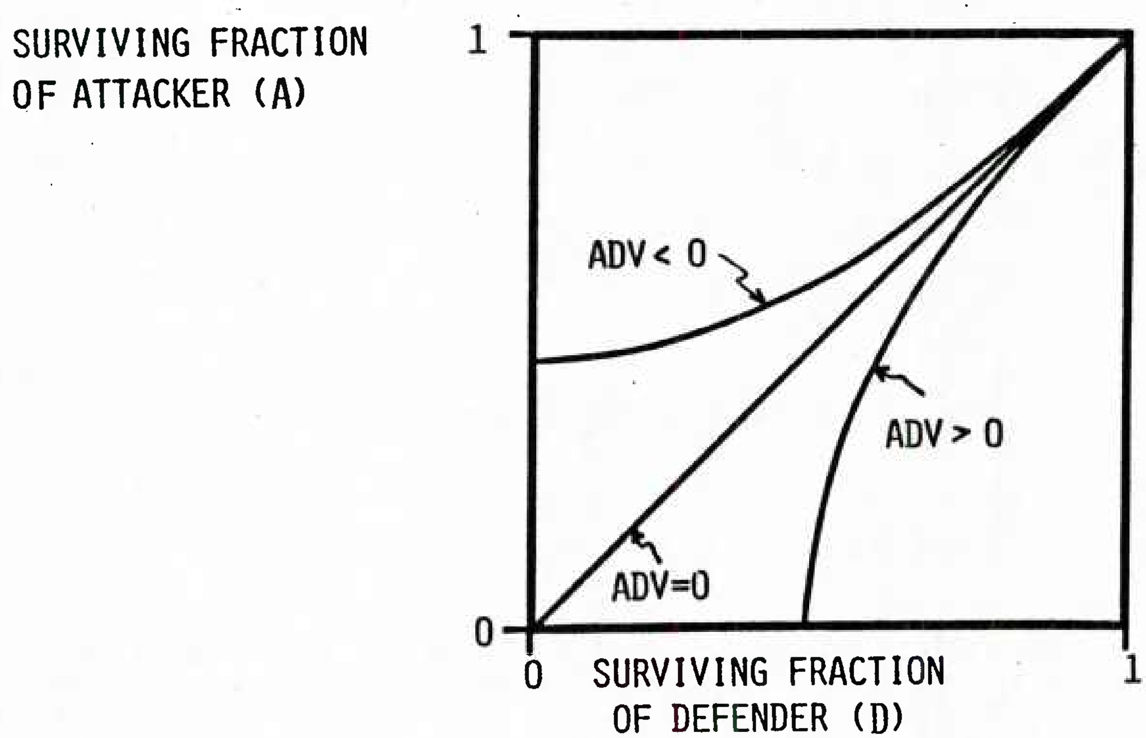


Figure 4-1. Effect of Advantage on Attrition History

d. Empirical Determination of ADV and EPS

(1) Empirical Formulas for MU, ADV and EPS. Equations (4-8), (4-5), (4-6) and (4-3) give the MU, EPS, and ADV parameters in terms of the Lanchesterian personnel activity parameters AA and DD. These formulas can, of course, be used only when the activity parameters are known a priori. However, **Refs 4-4 and 4-6** show that empirical estimates of MU, EPS and ADV can be obtained from empirical values of the initial and the final strengths, even though a priori values of the activity parameters are unknown. Now, the HERO data base does give the initial personnel strengths (XO and YO), and the personnel battle casualties (CX and CY) suffered in the course of the battle. The method of **Refs 4-4 and 4-6** sketched below shows how to use these data to obtain empirical estimates of MU, EPS and ADV. (Although the method obviously applies when X and Y are interpreted as empirical values for the surviving personnel at any time t after the start of the battle, most applications of it--including those in this paper--have to take t = T, i.e., they have to use the empirical values of X and Y at the end of the battle. The reason for this is, of course, that historical data are seldom available on surviving strengths at intermediate times during the battle.) When XO and YO are the initial personnel strengths, and CX and CY are the battle casualties at time t into the battle, the corresponding surviving strengths are

$$X = XO - CX$$

$$Y = YO - CY$$

and Equations (4-3) give the surviving personnel fractions as

$$A = X / XO$$

$$D = Y / YO$$

Then, as shown in **Refs 4-4 and 4-6,** Equations (4-4) can be solved for MU and EPS in terms of A and D to obtain the following empirical estimates of MU and EPS:

$$MU = SQR ((1 - A^2) / (1 - D^2)) \qquad (4\text{-}9)$$

$$EPS = LOG ((1 + MU) / (A + D * MU)) \qquad (4\text{-}10)$$

Equation (4-8) then yields the empirical estimate of ADV as

$$ADV = LOG (MU)$$

(2) Approximations to the Empirical Formulas for MU, EPS and ADV. For the battles in the computerized data base, EPS is often less than 0.2 or 0.3. The values of the hyperbolic functions for small values of EPS are shown in Table 4-1.

Table 4-1. Hyperbolic Functions for Small Values of EPS

EPS	COSH(EPS)	SINH(EPS)
0.0	1.00000	0.00000
0.1	1.00500	0.10017
0.2	1.02007	0.20134
0.3	1.04534	0.30452
0.4	1.08107	0.41075
0.5	1.12763	0.52110

From Table 4-1, we see that for sufficiently small values of EPS, the following approximations hold:

$$\mathrm{COSH(EPS)} = 1$$
$$\mathrm{SINH(EPS)} = \mathrm{EPS} \qquad (4\text{-}11)$$

Substituting these approximations into Equations (4-4), recalling that by definition FX = 1 - A and FY = 1 - D, and solving for MU and EPS yields the following approximations:

$$\mathrm{MU} = \mathrm{SQR\ (FX\ /\ FY)} = \mathrm{SQR\ (FER)} \qquad (4\text{-}12.1)$$

$$\mathrm{EPS} = \mathrm{SQR\ (FX * FY)}. \qquad (4\text{-}12.2)$$

Equations (4-12) will be called the linear approximations. By expanding the hyperbolic functions in a series and retaining in all calculations only terms of order EPS^3 or lower, the following more exact approximations can be derived:

$$\mathrm{MU}^2 = \mathrm{FER} * ((1 - \mathrm{FX}\ /\ 2)\ /\ (1 - \mathrm{FY}\ /\ 2)) \qquad (4\text{-}13.1)$$

$$\mathrm{EPS}^2 = (\mathrm{FX} * \mathrm{FY})\ /\ (1 - (\mathrm{FX} + \mathrm{FY})\ /\ 2) \qquad (4\text{-}13.2)$$

Equations (4-13) will be called the cubic approximations. To test the validity of these approximations, we compare the approximate values of MU (or of ADV = LOG(MU)) and EPS based on them to the exact values based on Equations (4-9) and (4-10). The results for the computerized data base are shown in Figures 4-2, 4-3 and 4-4, and can be summarized as follows:

(a) Figure 4-2 shows that ADV = LOG (MU) is approximately equal to (½) * LOG (FER), as asserted by Equation (4-12.1).

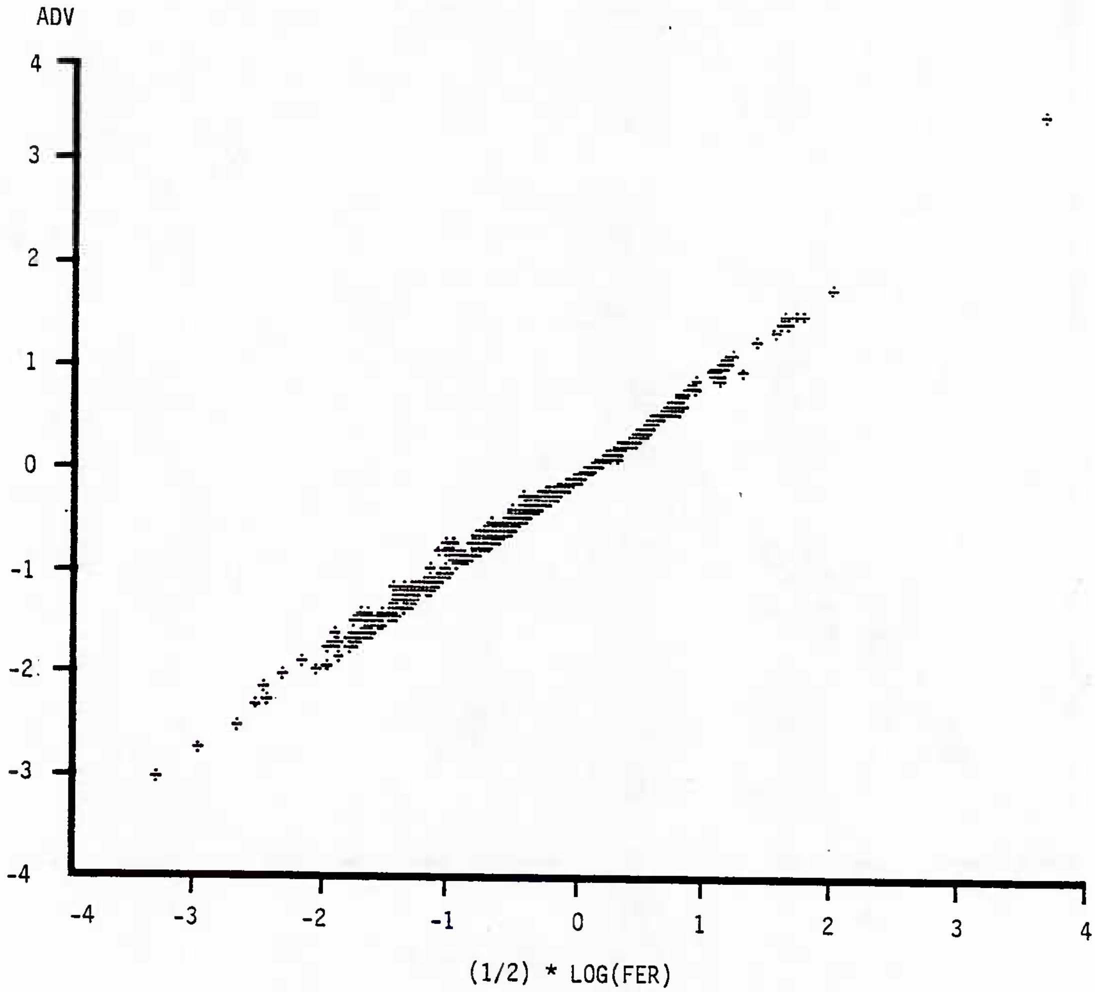


Figure 4-2. Comparison of Exact and Linear Approximation Values for ADV

(b) Equation (4-12.2) is a fairly good approximation when EPS is less than 0.2 (Figure 4-3).

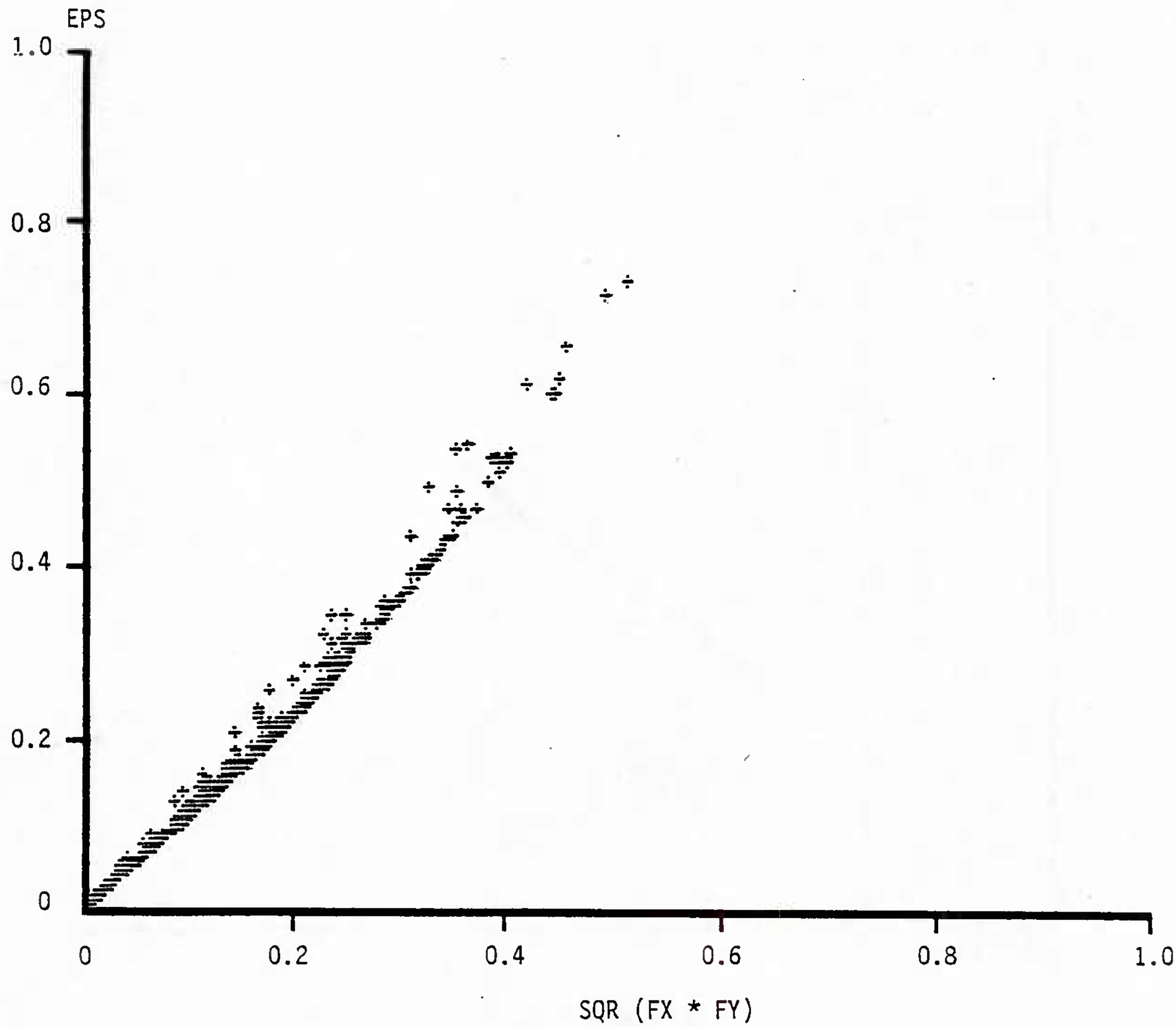


Figure 4-3. Comparison of Exact and Linear Approximation Values for EPS

Equation (4-13.2) is a better approximation, valid for the entire computerized data base (Figure 4-4).

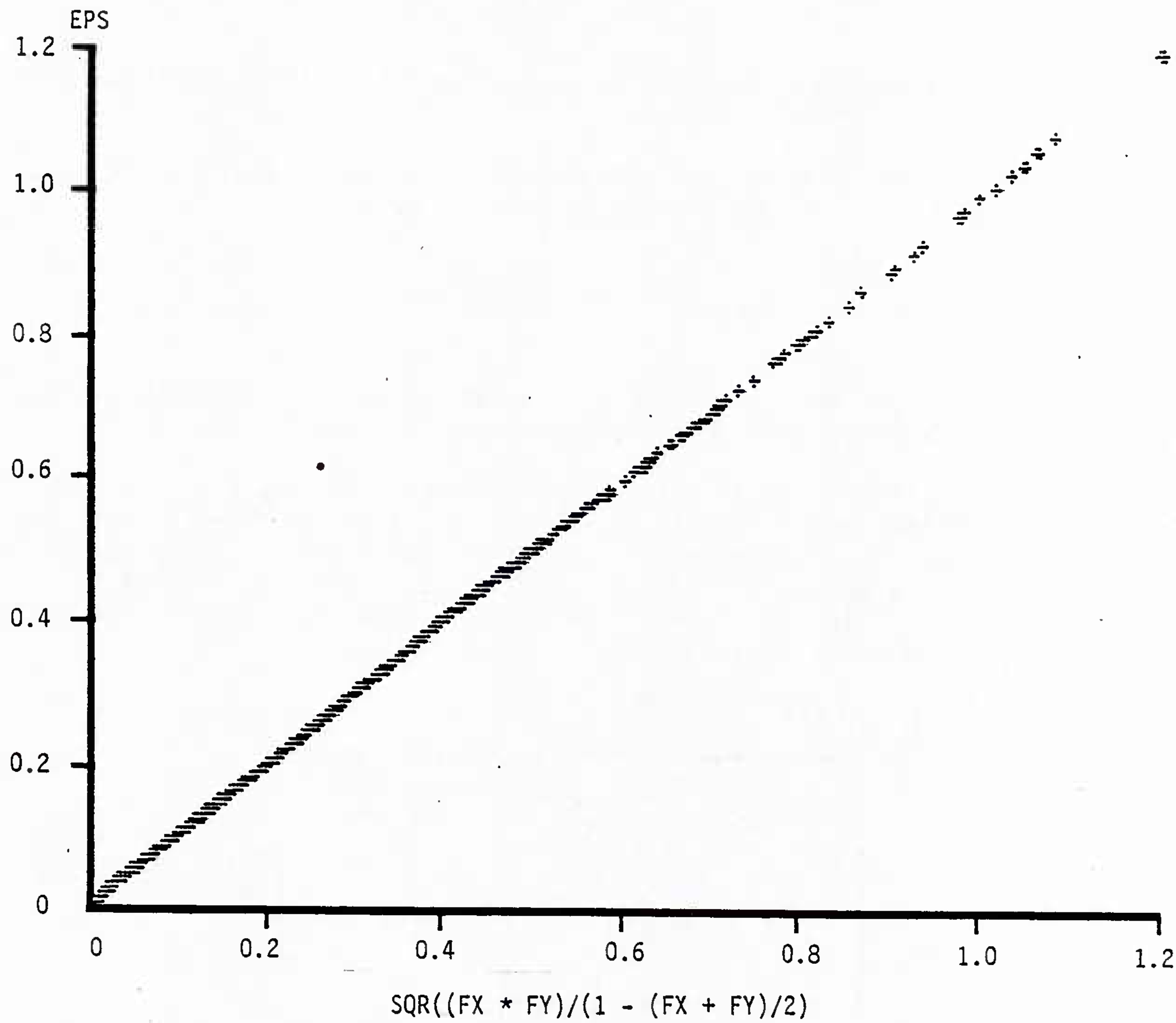


Figure 4-4. Comparison of Exact and Cubic Approximation Values for EPS

(3) **Interpretation.** These approximations illuminate the tactical significance of the parameters MU and EPS and confirm the theoretical interpretation of them offered in paragraph 4-2c. This is especially true for EPS (bitterness) since it is related directly to the geometric mean of the casualty fractions FX and FY as shown by Figure 4-3. Thus, EPS does indeed correspond to the nontechnical concept of the bitterness or bloodiness of a battle. The interpretation of ADV as an index of (the defender's) advantage is confirmed by Figure 4-5. That figure was generated by:

(a) Listing the battles in increasing order by their empirical ADV values,

(b) Segmenting this list into blocks of 40 contiguous battles each and averaging the ADV values for the battles in each block,

(c) Computing for each block the proportion of battles won by the attacker and the usual 95 percent confidence band about that proportion, and

(d) Plotting the values found in step (c) against those found in step (b), with a 95 percent confidence band on the proportion.

That the probability of an attacker victory depends strongly on ADV, and in particular declines precipitously as ADV changes from about -0.2 to +0.2, is beyond doubt. The method used to generate Figure 4-5 is technically crude and so has a number of serious limitations. However, this is a situation that is quite suitable for the application of logistic regression techniques, to which we will turn in paragraph 4-3.

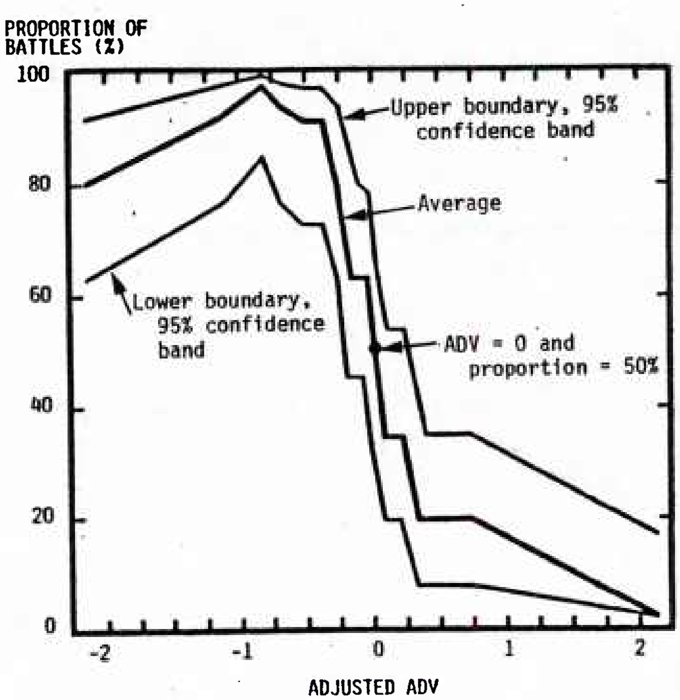


Figure 4-5. Proportion of Battles Won by Attacker versus Adjusted ADV

e. Determination of RESADV. References 4-4, 4-5, and 4-6 present evidence that, on the average, ADV depends approximately linearly on LOG(FR), so that:

$$ADV = a + b * LOG(FR) + RESADV, \quad (4\text{-}14)$$

where a and b are the so-called regression coefficients and the residual RESADV behaves like a normal random variable with zero mean. On the basis of empirical evidence, **Ref 4-4, 4-5, and 4-6** suggested that RESADV might be even more closely related to victory in battle than ADV. The empirical value of RESADV depends on what values are used for the regression coefficients, so we define the residual advantage relative to particular values of the regression coefficients to be:

$$RESADV(a,b) = ADV - a - b * LOG(FR). \quad (4\text{-}15)$$

RESADV(a,b) can be considered to be the residual value of ADV after the average effect of any differences in FR values is removed. **Reference 4-6** suggested on empirical grounds that the values a = 0 and b = -1/3 are fairly representative, so in this paper they are considered to be the "standard" values. Often RESADV(a,b) can be abbreviated to RESADV--usually the context will make it clear whether RESADV is to be interpreted as the general expression in Equation (4-15) or as the value relative to some particular choice of regression coefficients.

4-3. LOGISTIC REGRESSION

a. Orientation. Logistic regression techniques (see Appendix J) will be used to search for at least one variable that satisfies the criteria stated in paragraph 4-1. After reviewing the various logistic regression calculations that were considered, attention is focussed on the independent variables most closely associated with victory. The intimate association of these variables with victory is confirmed by a closer analysis and from several different points of view. Some observations are offered on the significance and application of these findings.

b. Choices for Logistic Regression Calculations. Many logistic regression calculations are conceivable, since the regression problem can be specified in various ways. All of the specifications addressed in this chapter are a subset of those outlined in Table 4-2, and the choices listed therein are explained later in this paragraph. Results and interpretations of the logistic regressions are presented in subsequent paragraphs.

Table 4-2. Choices for Logistic Regression Computations

1. Treatment of drawn battles

   1.1 Draws treated as draws, an outcome distinct from an attacker or a defender win

   1.2 Draws treated as a defeat to the attacker, and hence as a win for the defender

2. Data subsets

   2.1 All-HERO

   2.2 Pre-1940 or post-1940

   2.3 WWII or non-WWII

   2.4 1600-1699, 1700-1799, 1600-1799, 1800-1849, 1850-1899, 1900-1939, 1940-1949, 1950-1979

3. Independent variables

   3.1 ADV

   3.2 LOG(FER)

   3.3 RESADV

   3.4 LOG(CER)

   3.5 LOG(EPS)

   3.6 LOG(FR)

4. Strengths adjusted or unadjusted for replacement.

5. Symmetry forced or not forced.

(1) **Treatment of Draws.** In the data base, battle outcomes are recorded under WINA (see Glossary) as attacker wins, defender wins, or draws. It can be argued that draws should be lumped with the defender victories, since in drawn battles the defender stymies the attacker and prevents him from achieving his offensive ambitions. Although, for the most part, our logistic regression calculations treat draws as draws, in some cases the calculations were repeated with draws counted as defender wins in order to see how that would affect the results.

(2) **Data Subsets.** Various battle groupings can supply the observations to which the logistic functions are fitted. The battle groupings used in this chapter are indicated in Table 4-2.

(3) **Independent Variables.** In this paper, each of the variables identified in paragraph 4-1b and repeated in Table 4-2 were used as the independent variable in one or more logistic regression calculations. Of course, considering the findings of paragraph 4-2, we anticipate that:

(a) Using ADV or LOG(FER) as the independent variable should lead to essentially the same logistic regression results. By Equations (4-12.1) and (4-8), we have the linear approximation:

$$ADV = LOG(MU) = (\tfrac{1}{2}) * LOG(FER),$$

so that ADV is approximately half LOG(FER).

(b) LOG(EPS) should be only weakly related to WINA, since by paragraph 4-2c EPS theoretically does not affect winning or losing. The logistic regression results presented later (see Table 4-3) tend to confirm these expectations.

(4) **Adjustment of Strengths.** Paragraph 4-2 defines the independent variables in terms of the initial and final personnel strengths of the engaged sides in a battle. But the data base gives "total engaged" personnel strengths which for most of the battles are the desired initial strengths, but which for some battles are either average daily strengths or total strength committed during the course of the battle. Unfortunately, the HERO data base does not identify which "total engaged" values are initial and which are not. Clarification of this situation is part of the CDES contract, as explained in Appendix I (paragraph I-3c) but the results were not available for use in this paper. Accordingly, some of the logistic regression calculations use the "total engaged" values as though they were in all cases the initial strengths--these are called the unadjusted strengths. However, in most of the logistic regression calculations, the following procedure was used to adjust the "total engaged" values to approximate the effect of replacements:

(a) If the battle duration T is less than 10 days, the initial strength is taken equal to the "total engaged" strength.

(b) If the battle duration is at least 10 but less than 20 days, the initial strengths are taken to be:

XO = Total Engaged (ATK) + CX /2

YO = Total Engaged (DEF) + CY /2

(c) If the battle lasts 20 days or more, the initial strengths are taken to be:

XO = Total Engaged (ATK) + CX

YO = Total Engaged (DEF) + CY.

(d) In all cases, final strengths are calculated as:

X = XO - CX

Y = YO - CY.

This adjustment process is clearly only a rough approximation to the effects of replacements over a lengthy battle. Fortunately, this chapter's logistic regression results are nearly the same whether adjusted or unadjusted strengths are used. This is partly due to the fact that battles in the HERO data base seldom continue for as long as 10 or 20 days. For example, only about 4 percent of the battles lasted at least 10 but less than 20 days. Another 4 percent lasted 20 days or more (see, for example, the columns labeled "Empirical" in Chapter 3, Table 3-4).

(5) Symmetry. In the notation of Appendix J, a logistic function is said to be symmetric if

$$\Pr(\underline{x}_n) = 1 / (1 + R)$$

for all n = 1(1)N whenever $x_{np} = 0$ for n = 1(1)N and p = 1(1)P. The logistic function fitted to the observations can be <u>forced</u> to be symmetric simply by setting $x_{n0} = 0$ for n = 1(1)N. On the other hand, if $x_{n0} = 1$ for n = 1(1)N, then symmetry is <u>not</u> forced and the fitted logistic function may or may not turn out to be symmetric. Symmetry was forced in the numerical example of Appendix J, paragraph J-5. However, for that example, the fitted function would be symmetric in any case because the observations are symmetric (in the sense of reflection through the point at x = 0 and $P_1(0)$ = 50 percent, as shown in Appendix J, Figure J-1). For most of the logistic regression calculations in this chapter, symmetry is not forced, but in some instances a close approximation to it arises naturally from the fitting process.

c. Logistic Regression Findings

(1) Selection of Variables for Further Analysis. The selection of variables for detailed investigation will be done by choosing, from among the six variables in Table 4-2, those that best fit the data on battle outcomes for the non-WWII data subset. The situation for the WWII data subset will be addressed in paragraph 4-4. Here, draws are counted as draws, strengths are adjusted, and symmetry is not forced. The basic results of the logistic regression computations for this situation are presented in Table 4-3. The column labeled L(0) gives the loglikelihood value when all of the fitted parameters are set equal to zero (cf. Appendix J, Equation (J-14)). The column labeled MAX.L gives the maximum loglikelihood value reached by the DALOFIT logistic regression program. The columns labeled a(1,0), a(1,1), a(2,0), and a(2,1) give the maximum likelihood parameter values of the logistic function fitted to the data subset used. Here a(r,p) is the logistic regression coefficient for essential response level r and parameter p, with r=1 used for a draw and r=2 used for ATK wins. The columns labeled SD(1,0), SD(1,1), SD(2,0), and SD(2,1) give the standard deviations of the maximum likelihood parameters. Thus, SD(1,0) is the estimated standard deviation of a(1,0), etc.

Table 4-3. Logistic Regression Results[a]

Independent variable	Number of data points	L(0)	MAX.L	a(1,0)	SD(1,0)	a(1,1)	SD(1,1)	a(2,0)	SD(2,0)	a(2,1)	SD(2,1)
ADV	427	-469	-219	-1.527	0.26	-3.783	0.80	0.247	0.15	-5.997	0.63
LOG(FER)	427	-469	-219	-1.522	0.26	-1.733	0.37	0.242	0.15	-2.770	0.29
RESADV[b]	427	-469	-222	-1.214	0.24	-3.477	0.78	0.770	0.16	-6.136	0.63
LOG(CER)	427	-469	-239	-1.248	0.26	-1.225	0.32	0.888	0.16	-2.308	0.24
LOG(EPS)	427	-469	-354	-1.832	0.54	0.013	0.22	0.905	0.27	0.164	0.11
LOG(FR)	435[c]	-478	-362	-1.892	0.25	0.364	0.30	0.468	0.11	0.326	0.16

[a]For draws counted as draws, non-WWII data subset, adjusted strengths, and symmetry not forced.

[b]The standard values RESADV(0,-1/3) are used.

[c]Eight non-WWII battles have data for both X0 and Y0, but are missing data on either CX or CY.

(2) **Ranking of Variables.** A rough measure of the relative quality of the logistic regression fits is provided by the increase in loglikelihood, i.e., by the quantity:

$$\text{MAX.L} - \text{L}(0).$$

For this measure, it is seen that the variables ADV, LOG(FER), and RESADV are approximately tied for best fit. The variable LOG(CER) is fourth best. The variables LOG(EPS) and LOG(FR) are approximately tied for worst fit. Table 4-3 also shows that the variables ADV and LOG(FER) are essentially equivalent with regard to logistic regression as can be seen from the facts that:

(a) The fitted parameters a(1,0) and a(2,0) for LOG(FER) are practically the same as for ADV. The corresponding standard deviations SD(1,0) and SD(2,0) are also practically the same.

(b) The fitted parameters a(1,1) and a(2,1) for LOG(FER) are approximately half those for ADV--as expected from the fact that LOG(FER) is approximately twice ADV, as was shown in paragraph 4-3a(3). The corresponding standard deviations SD(1,1) and SD(2,1) also follow this pattern.

d. ADV and Probability of Victory

(1) Fitted Logistic Functions. The logistic functions fitted to the non-WWII data subset are plotted in Figure 4-6. While Figure 4-6 is conceptually similar to Figure 4-5, it provides a much better and more detailed view of the connection between ADV and battle outcome.

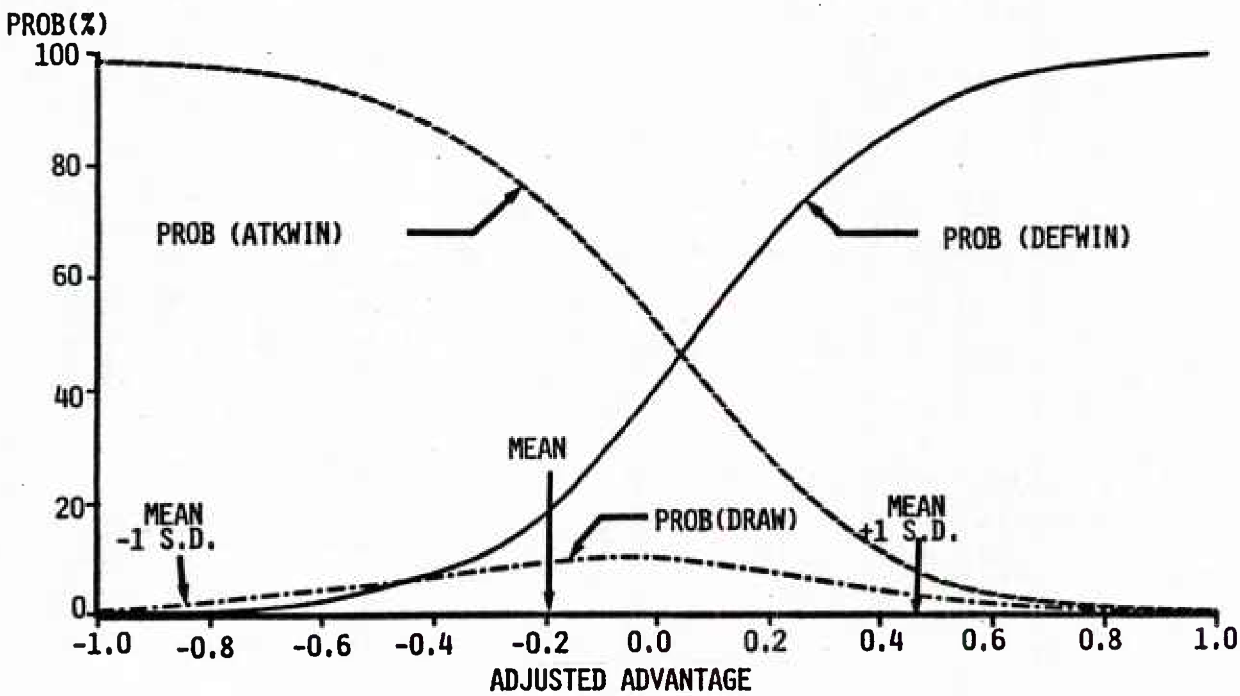


Figure 4-6. Probability of Battle Outcome for Non-WWII Battles versus Adjusted Advantage

Figure 4-6 shows that the defender's probability of victory rises sharply as ADV increases. Also, PROB(DRAW) rises to a maximum near ADV = 0, and at that point PROB(ATKWIN) is about equal to PROB(DEFWIN), again confirming that ADV is a measure of the defender's advantage--more drawn battles occur when ADV = 0 because the two sides are about evenly balanced. Although symmetry was not forced, the curves for PROB(DEFWIN) and PROB(ATKWIN) are nevertheless nearly symmetric. When drawn battles are lumped with defender wins, the curves for PROB(DEFWIN) and PROB(ATKWIN) are almost exactly symmetric. The attacker won the greater proportion of non-WWII battles, and in fact for this data subset ADV tends to be negative (so that the defender was at a disadvantage in most of the battles). This is shown in

Figure 4-6 by the arrows designating the MEAN value of ADV, the MEAN + 1 SD, or the MEAN - 1 SD, the MEAN and SD being for the adjusted ADV values in the non-WWII data subset. That the greater proportion of attacker victories is reflected in a tendency toward lower ADV values, rather than in an asymmetry of the curves for PROB(DEFWIN) and PROB(ATKWIN), is further evidence that ADV has a very deep and fundamental connection with victory in battle. Thus, ADV appears to have the sought-for properties listed in paragraph 4-1. The other variables tied with ADV for best fit also possess the sought-for properties. However, the theoretical rationale for the relation of ADV (or equivalently of LOG(FER)) to victory is currently stronger than for RESADV. For this reason, the remainder of this chapter focuses on ADV and LOG(FER) as the variables most closely associated with victory in battle. Additional important information about them will be developed in subsequent paragraphs of this chapter.

(2) **Observed and Fitted Probabilities of Victory.** A key issue is whether the fitted logistic functions give the correct probability of victory. A plot of the observed versus the fitted probability of victory provides a visual representation of the fit. Figure 4-7 shows a plot of this type for the non-WWII data when adjusted ADV is used as the independent variable in the logistic regression function. The fit to the probability that the attacker wins is very good, as shown by the fact that the observed proportions of attacker victories generally fall close to their fitted values.

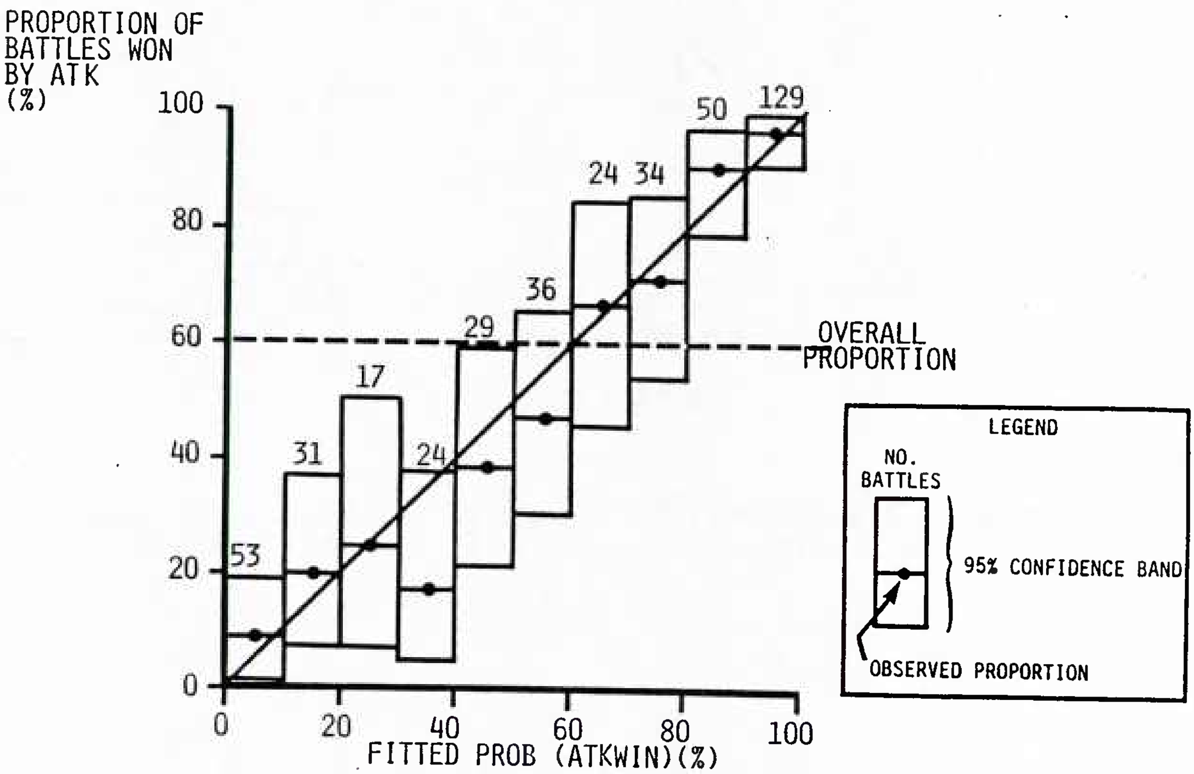


Figure 4-7. Proportion of Battles Won by Attacker and Fit Based on Adjusted ADV for Non-WWII Data (427 battles)

Figure 4-8 shows a plot of the observed versus the fitted probability of an attacker victory based on adjusted LOG(FR). It reveals that the fitted probability of an attacker victory nearly always predicts win probabilities close to the average overall proportion of attacker victories. So the logistic regression fit based on LOG(FR) does not identify those battles whose probability of attacker victory is markedly higher or lower than the average. Hence, one could do almost as well simply by using the average proportion of attacker victories as by using the fitted probability. Accordingly, LOG(FR) is not nearly as precise a determiner of victory as is ADV.

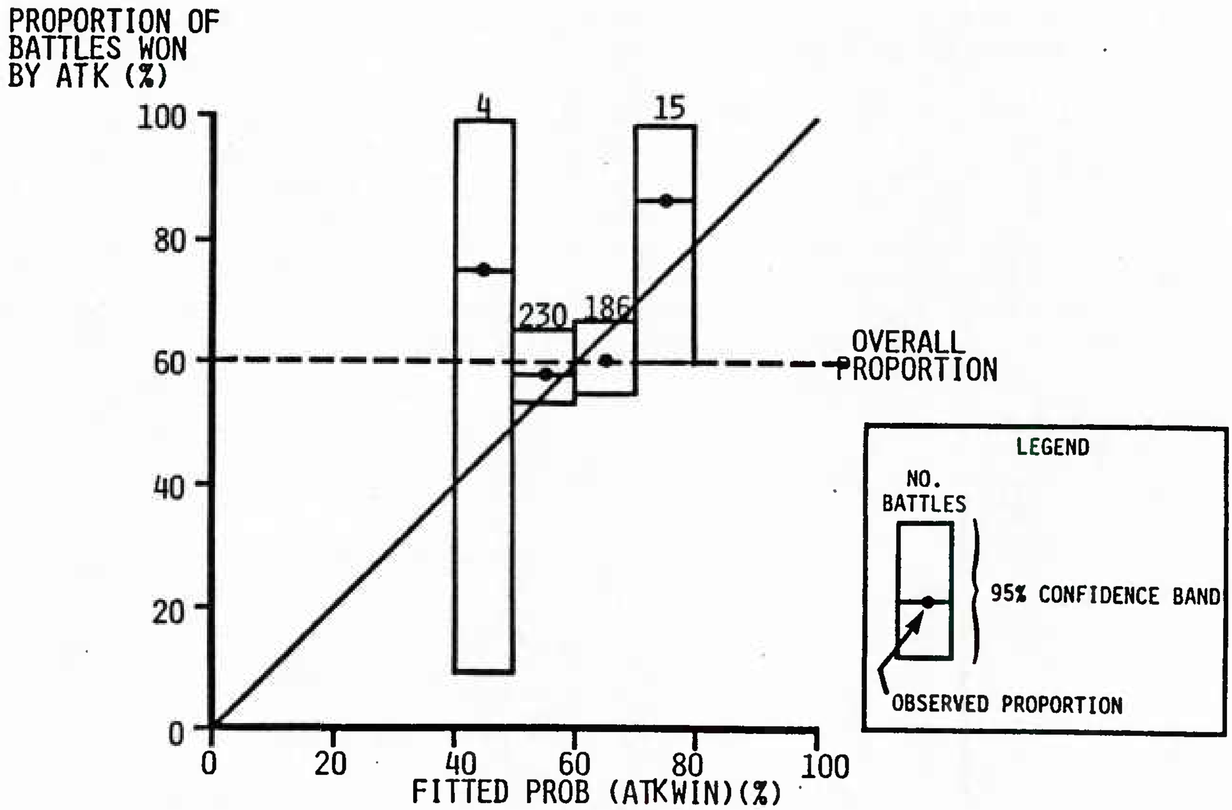


Figure 4-8. Proportion of Battles Won by Attacker and Fit Based on Adjusted Force Ratio for Non-WWII Data (435 battles)

Figure 4-9 shows how the observed probability of an attacker victory in the CORG data base of 175 battles compares with those predicted using the logistic function fitted to the non-WWII subset of the HERO data base. Although the observed proportion of attacker victories seems to be somewhat higher than expected for fitted probabilities of 30 percent or less, the overall agreement is acceptable. This indicates that the logistic functions fitted to the non-WWII subset of the HERO data can be applied successfully to other data bases.

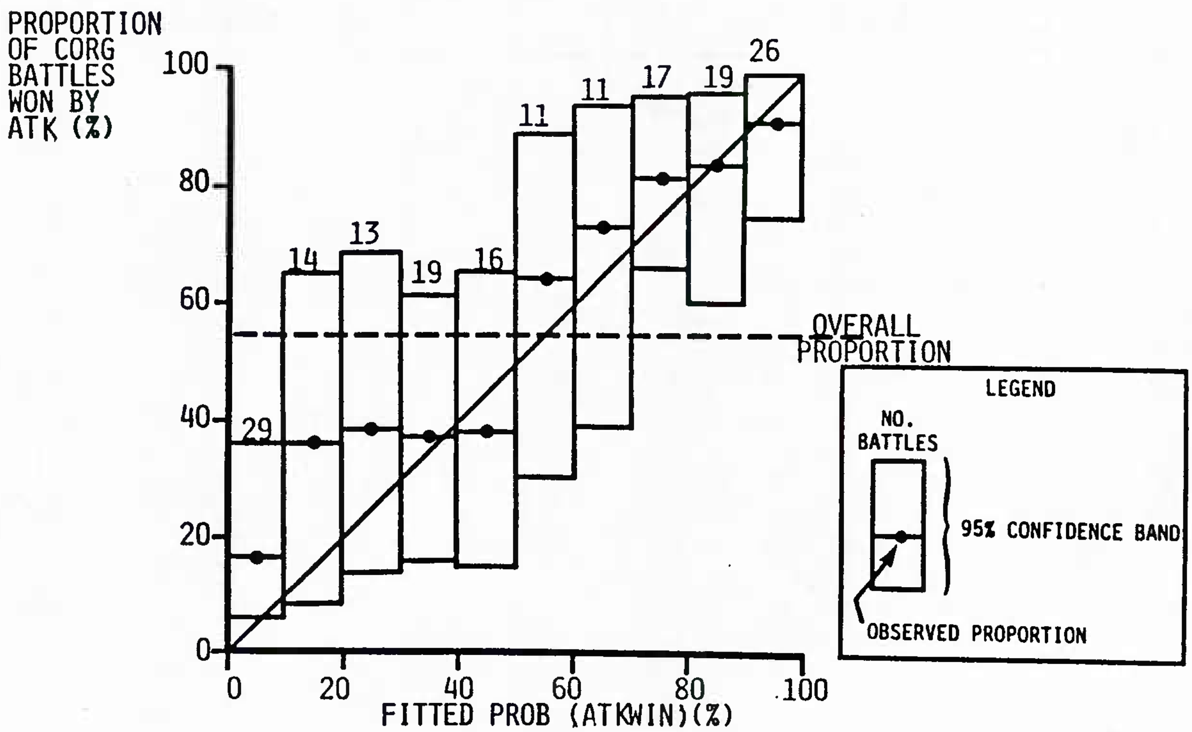


Figure 4-9. Proportion of CORG Data Base Battles Won by Attacker and Fit Based on Adjusted ADV for Non-WWII HERO Battles

Table 4-4 shows how this works out for still another data set--one specifically chosen to contain a high number of battles that occurred either very early or very late in time. None of these battles appear in the non-WWII HERO data subset. The degree of agreement is very encouraging, and suggests that the relation between ADV and victory in battle has persisted essentially unchanged for a remarkably long period of time. Because of this persistence, it is reasonable to expect it to persist for the foreseeable future. This further confirms the choice of ADV and LOG(FER) as the variables to subject to further analysis.

Table 4-4. Predicted and Observed Winner for Some Battles of Extreme Dates

No	Date	Name	Observed ADV	Predicted[a] P(ATKWIN)	Reported winner
1	1944	Kwajalein North	-1.30	0.99	ATK
2	1944	Kwajalein South	-1.10	0.98	ATK
3	1944	Eniwetok	-1.00	0.98	ATK
4	1222	Indus	-0.95	0.98	ATK
5	1512	Ravenna	-0.61	0.94	ATK
6	1943	Attu	-0.60	0.94	ATK
7	1944	Guam	-0.53	0.92	ATK
8	1944	Saipan	-0.42	0.89	ATK
9	1945	Iwo Jima	-0.36	0.86	ATK
10	1982	Falkland Islands	-0.2 to -0.9	0.74 to 0.99	ATK
11	280 B.C.	Heraclea	-0.18	0.73	ATK
12	1562	Dreux	-0.13	0.67	ATK
13	1968	Khe Sanh	0.16	0.31	DEF
14	351	Mursa	0.18	0.28	ATK
15	1515	Marignano	0.30	0.16	DEF
16	279 B.C.	Asculum	0.33	0.14	DEF
17	1386	Sempach	0.52	0.05	DEF
18	1944	Driniumor River	0.82	0.01	DEF

[a]Prediction using observed ADV and fit to non-WWII data base.

(3) Observations

(a) On the Relation of Victory and ADV. In view of the theoretical interpretations offered in paragraph 4-2c, the findings: (1) that ADV and LOG(FER) are essentially equivalent, and (2) that they measure the defender's advantage can be explained by postulating that forces engaged in battle are "rational" in the sense that they have a very strong tendency to get out of the situation when the ADV or LOG(FER) values are unfavorable to them. Thus, a side that loses 10 percent of its personnel while its opponent loses 15 percent sees that its opponent is weakening faster than it is, and so rationally should continue to fight. The opponent, on the other hand, is anxious to break off the engagement so he can try to find a more favorable situation. To the extent that this is what really happens in battles, the conventional "breakpoint" methods for ending simulated battles may be badly in error because they fail to allow the termination to depend on FER (cf. Chapter 6). Analogously, rates of advance against enemy opposition may be found to depend much more on FER than on FR. The opposing forces may be able to sense their ADV values, for according to Clausewitz, "Usually, a battle takes shape from the start, though not in any obvious manner. Often this shape has already been decisively determined by the preliminary dispositions made for the battle, and then it shows lack of insight in the commander who opens the engagement under these unfavorable conditions without being aware of them. Even if the course of the battle is not predetermined, it is in the nature of things that it consists in a slowly shifting balance, which starts early, but, as we have said, is not easily detectable. As time goes on, it gathers momentum and becomes more obvious. . . . But . . . it is certain that a commander usually knows that he is losing the battle long before he orders retreat. Battles in which one unexpected factor has a major effect on the course of the whole usually exist only in stories told by people who want to explain away their defeats." (**Ref 4-7**, page 249)

(b) On the Relation of ADV to Other Factors. Note that when we use logistic regression with ADV as the independent variable we have thrown away--or at any rate have made no direct use of--information on such other factors as:

(1) Battle date

(2) Locale or terrain

(3) Weather

(4) Morale

(5) Training

(6) Tactical plans or maneuvers by the attacker or by the defender

(7) Logistics

(8) Surprise

(9) Fortifications

(10) Battle duration

(11) Bitterness or intensity

(12) Force mixes (such as cavalry, tanks, artillery, or air)

(13) Etcetera

In fact, ADV does not even make any direct use of the initial force strengths. It uses directly only the information contained in the values

$$FX = CX \ / \ XO, \text{ and}$$

$$FY = CY \ / \ YO.$$

Moreover, even these are telescoped into a single index (MU or ADV) via Equations (4-9) and (4-8). In view of the frequency with which other factors are mentioned as the causes of victory, it may be surprising that ADV--and FER--are so intimately related to victory in battle. Yet the connection of ADV (or FER) with victory in battle seems to be a very deep and fundamental one that holds, on the average, despite all sorts of variations in tactics, force mixes, weather, terrain, morale, leadership, surprise, logistical support, training, technology, force ratios, etc. These findings can be explained if we postulate that the influences of all these other factors on victory are captured in or expressed by the ADV or FER. That is, we postulate that ADV has a direct connection with victory in battle, while the other factors have only an indirect effect on victory. The postulated causative sequence is as follows:

1. Factors such as chance, accidents, morale, leadership, logistics, etc. directly influence personnel losses.

2. Personnel losses directly influence FX and FY.

3. FX and FY directly influence FER and ADV.

4. ADV and LOG(FER) directly influence victory.

Presumably, forces gradually become aware of the effects of a favorable or adverse FER or ADV as the battle progresses. If we also postulate that forces have difficulty in sensing whether their ADV is favorable or unfavorable when their ADV is close to zero--but can sense it more easily when it is very high or very low--then we can derive the following inference, which is in principle testable by appeal to the data:

- Battles with ADV values near zero tend to be more bitter, take longer, and are more likely to lead to draws than battles with very high or very low ADV values, and if not drawn are about equally likely to be won by either side.

Another interesting conjecture is that victory depends exactly on ADV, i.e., that the curve of victory versus ADV in Figure 4-6 is <u>theoretically</u> a "step function" with zero probability of defender victory for negative ADV values and unit probability of defender victory for positive ADV values. Explanations why the <u>observed</u> curve for P(DEFWIN) rises smoothly as ADV increases, rather than being a step function, include the following:

- The engaged sides only inaccurately perceive the true value of ADV or FER.

- The engaged sides often can react only sluggishly to a perceived ADV value--they are often unable to seize an advantage quickly enough to press it home, and are unable to extricate themselves from an unfavorable situation quickly enough to avoid suffering more casualties than they should.

- The values of ADV and FER fluctuate somewhat during a battle, thus clouding each side's perception of the situation.

- Although forces may realize their situation with respect to ADV, they choose not to respond rationally to it because they do not realize how closely associated it is with victory, because they are victims of a sort of wishful thinking that in spite of current conditions things will get better, or because conditions beyond the scope of the immediate battlefield require either a more strenuous defense or a more cautious attack than would be the case were external considerations not a factor.

- Some of the data may incorrectly award victory to the side that lost the battle.

- Some of the strength and loss data are inaccurate.

(c) **ADV Should Be Used as a Payoff Function.** Since the curves for PROB(DEFWIN) and PROB(ATKWIN) are nearly symmetric, each side can increase its relative advantage only at the expense of decreasing by the same amount its opponent's. Thus, each side seems to be in a zero-sum game with either ADV or FER as the payoff function that each is striving to optimize (the defender is trying to increase it, and the attacker is trying to decrease it). Accordingly, ADV should be used in studies and analyses as the payoff function or figure of merit for assessing the value of alternative organizations, tactics, equipment, and force mixes. Soldiers and commanders should be taught in their service schools, academies, war colleges, and staff colleges that high values of FER are strongly associated with winning battles--and therefore that increasing, or even appraising, the value of their FER could be very important in battles and similar tactical engagements. Perhaps computation of ADV or FER during the early stages of a battle would improve tactical decisions for the conduct of the rest of the battle. If at an early stage, the FER value is found to be unfavorable, then the commander should either immediately seek additional support or other means for improving his FER, or else he should attempt to break off the engagement as expeditiously as possible and to find more favorable circumstances for engaging the enemy.

(d) **Use of ADV for Historical Analysis and Rating of Forces.** The relation of ADV to victory in battle can be used for historical criticism and analysis. For example, if a force that had a large probability of winning the battle reportedly lost it, this is sufficient reason to review the evidence more closely to determine whether the historical reports are accurate and, if they are, what caused this unexpected and unusual turn of events. ADV or FER may also be used to rate the performance of historical captains--commanders that were consistently able to achieve favorable FER or ADV values would rate highly. A similar rating system for friendly units in time of war may be possible--provided, of course, accurate and reliable data on friendly and enemy forces and losses are available.

(e) **Simulating a Commander's Level of Confidence.** The relation of ADV to victory in battle could be used in war games to simulate a commander's level of confidence in winning a battle. A specific application of this idea to escalation from conventional to tactical nuclear or chemical usage has been proposed in **Ref 4-6**, to which the reader is directed for more details.

(f) **Testing War Simulations and Theories of Combat.** Moreover, the relation of ADV and FER to victory in battle can be used to test wargames and theories of combat for realism. If the wargame or theory of combat determines a probability of victory that is inconsistent with the empirically observed relationship of ADV to victory, then that wargame or theory of combat is highly suspect and its results should be used with extreme caution.

4-4. THE WORLD WAR II ANOMALY

a. **Orientation.** Paragraph 4-3 focussed on choosing a variable that is closely associated with victory, using mainly the non-WWII data subset. That data subset was used because the WWII data appear to be anomalous. This paragraph describes the WWII anomaly and presents the results of some attempts to identify its source. Suggestions on further steps for analyzing the WWII anomaly are discussed in paragraph 4-5, below.

b. Changes in Logistic Regression Results Over the Years. Logistic regression calculations using adjusted ADV as the independent variable were done for each of the data subsets listed in Table 4-2. The results of these logistic regressions are exhibited in Table 4-5.

Table 4-5. Selected Logistic Regression Results[a]

Data subset	Number of data points	L(0)	MAX.L	a(1,0)	SD(1,0)	a(1,1)	SD(1,1)	a(2,0)	SD(2,0)	a(2,1)	SD(2,1)
All HERO	585	-643	-380	-1.82	0.20	-1.31	0.41	0.247	0.11	-2.68	0.26
Pre-1940[b]	374	-411	-186	-1.41	0.26	-3.35	0.88	0.311	0.17	-6.61	0.74
Post-1940[b]	211	-232	-154	-1.90	0.40	-0.472	0.56	0.518	0.18	-0.990	0.27
Non-WWII[b]	427	-469	-219	-1.53	0.26	-3.78	0.80	0.247	0.15	-6.00	0.63
WWII[b]	158	-174	-116	-1.77	0.43	0.314	0.62	0.624	0.21	-0.613	0.28
1600-1699	46	-51	-8	-6.26	17	-2.22	34	2.20	0.90	-6.87	2.4
1700-1799	65	-71	-32	-2.87	1.0	-1.72	2.7	0.416	0.34	-4.30	1.2
1600-1799	111	-122	-43	-3.05	1.1	-2.16	2.6	0.804	0.30	-4.73	1.0
1800-1849	51	-56	-22	-0.451	0.71	-6.46	3.4	0.679	0.59	-12.1	3.7
1850-1899	74	-81	-37	-1.97	0.62	-2.60	2.1	-0.0770	0.35	-6.11	1.5
1900-1939	138	-152	-73	-0.944	0.35	-4.231	1.5	-0.0101	0.30	-8.52	1.6
1940-1949	158	-174	-116	-1.77	0.43	0.314	0.62	0.624	0.21	-0.613	0.28
1950-1979	53	-58	-27	-3.85	1.4	-6.96	2.2	-0.262	0.55	-5.27	1.7

[a]For adjusted ADV as the independent variable, draws counted as draws, adjusted strengths, and symmetry not forced.

[b]Pre-1940 includes the years 1600-1939 (inclusive).
Post-1940 includes the years 1940-1979 (inclusive).
WWII includes the years 1940-1949 (inclusive).
Non-WWII includes the years 1600-1939 and 1950-1979 (inclusive).

Figure 4-10 shows the fitted values for the probability that the attacker wins versus adjusted ADV for the all-HERO, pre-1940, and post-1940 data subsets. Visual inspection of Figure 4-10 suggests that the curves for the pre-1940 and post-1940 subsets may have significantly different shapes. Since the shapes of these curves are largely controlled by the logistic regression parameter a(2,1), defined in Appendix J, it can be used to help investigate suspected differences in shape. For example, the value of a(2,1) for the pre-1940 data subset, plus or minus two standard deviations, yields a confidence band of -8.1 to -5.1. A similar plus or minus two standard deviations confidence band on a(2,1) for the post-1940 data subset runs from -1.5 to -0.5. Since there is a relatively wide gap separating these two confidence bands, it is reasonable to conclude that the post-1940 data subset differs statistically from the pre-1940 data subset with regard to the dependence of victory in battle on ADV. The fact that the post-1940 subset is anomalous is referred to as the WWII anomaly because it starts with World War II and because, as we shall see below, the WWII subset is a major contributor to this anomaly.

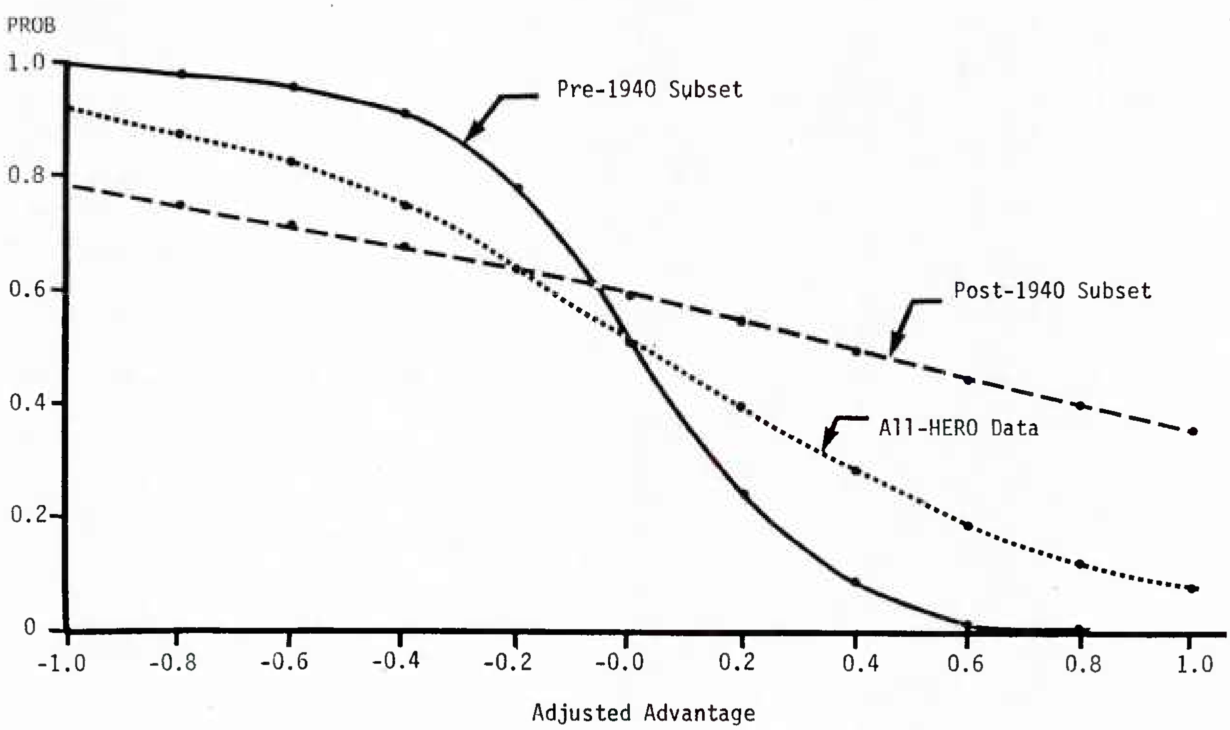


Figure 4-10. Probability of Battle Outcome Versus Adjusted Advantage for the All-HERO, Pre-1940, and Post-1940 Data Subsets

c. First Attempts to Localize the Source of the World War II Anomaly. For the first attempt to localize the source of the WWII anomaly, the data were grouped into subsets by battle date, making an effort to keep the number of battles in each subset large enough to retain some stability in the logistic regression fits--which meant that subsets with fewer than 50 battles were avoided as much as possible, and that subsets with at least 100 battles were preferred. The subsets that were used are as indicated in Tables 4-2 and 4-5. Figure 4-11 shows the fitted probability of an attacker victory versus adjusted ADV for several of these data subsets. Visual inspection of these curves suggests that, with the exception of the World War II decade of 1940-1949, the relation between victory in battle and ADV has not changed much over time. Inspection of Table 4-4 tends strongly to confirm this stability. Thus, the anomalous logistic regression results appear to be associated mainly with the World War II data subset.

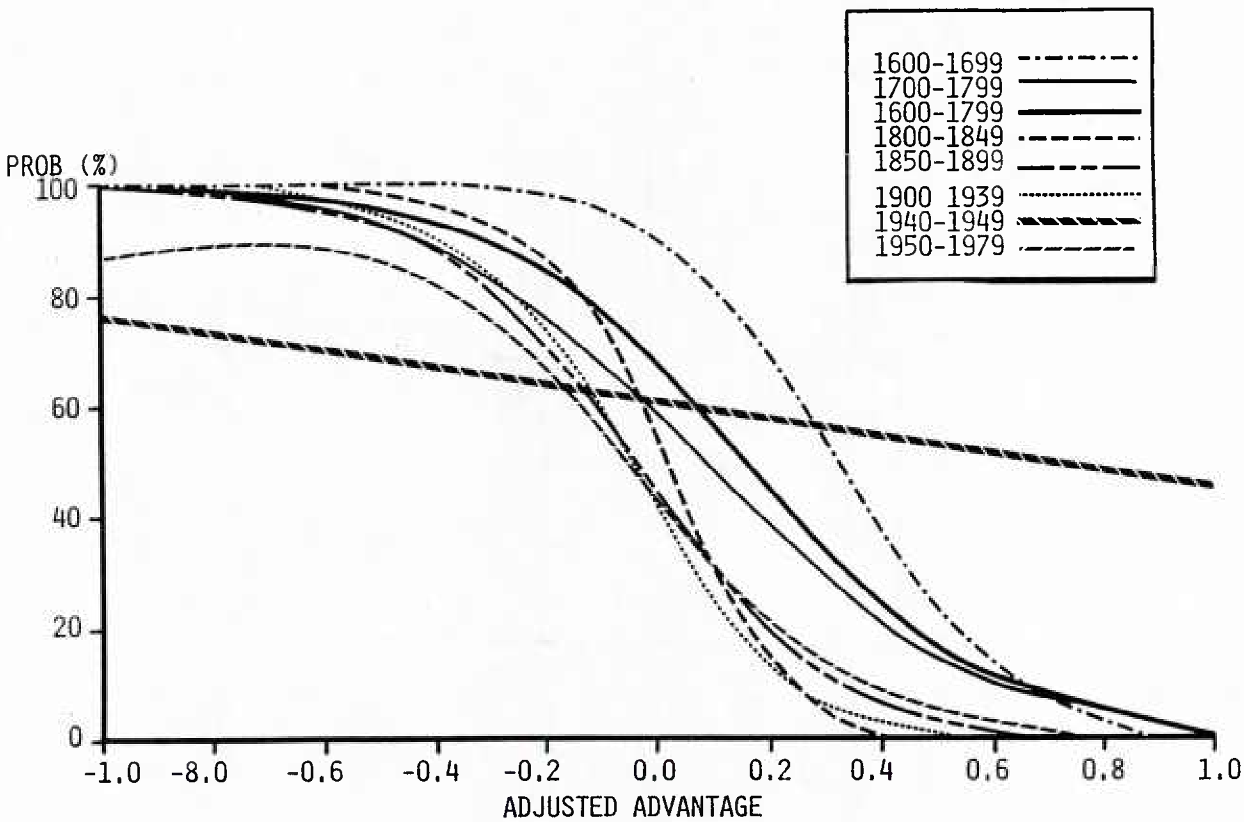


Figure 4-11. Probability Attacker Wins Versus Adjusted Advantage for Selected Time Periods

Figure 4-12 plots the values of the logistic regression parameter a(2,1) with their plus or minus two standard deviation confidence bands. It shows that the World War II data subset is quite different from the other data subsets, all of which have confidence bands that overlap the likely zone of a(2,1) values for the non-WWII data subset.

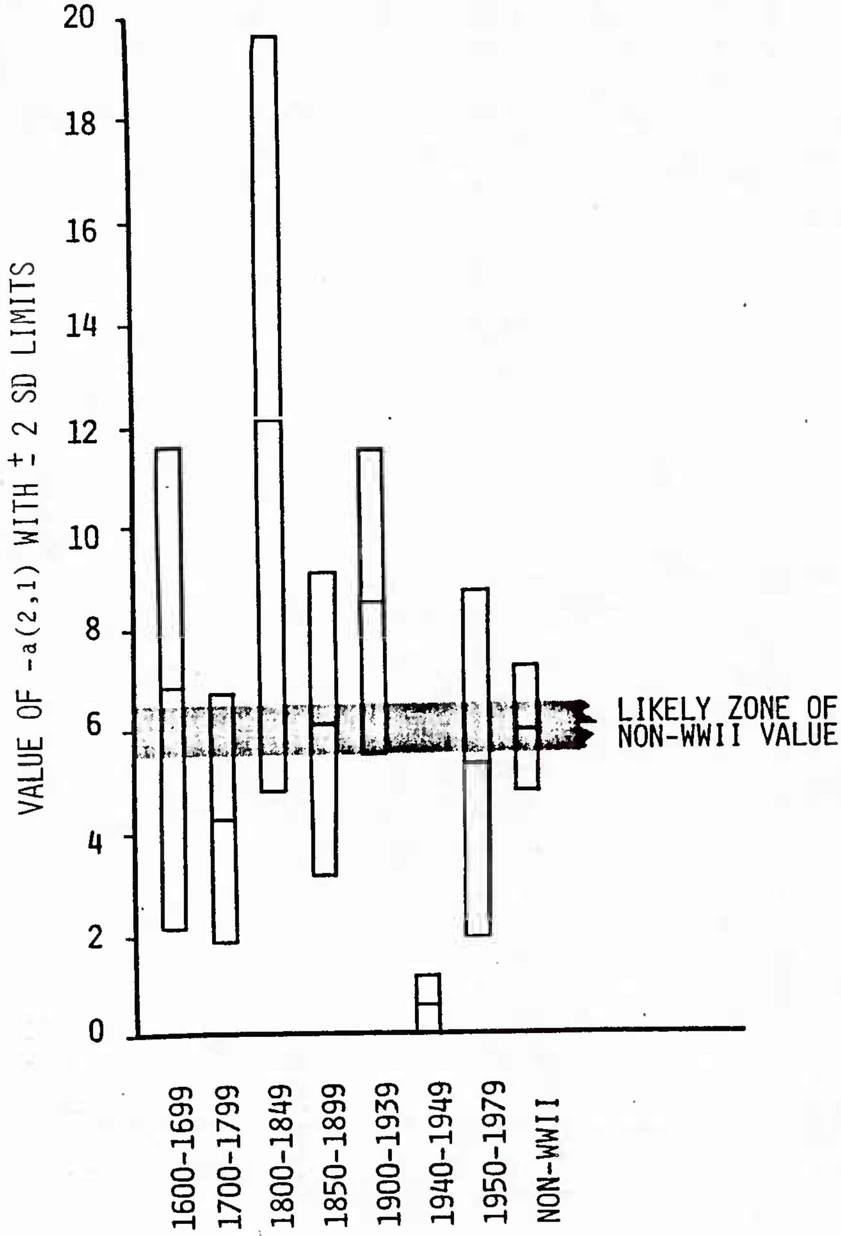


Figure 4-12. Mean and Two-Standard Deviation Confidence Bands for the Logistic Regression Parameter a(2,1)

Figure 4-13 illustrates that the WWII data subset differs from the others with respect to its logistic regression parameter a(2,1), but not with respect to its logistic regression parameter a(2,0). In the following paragraphs we will seek to further localize the source of the World War II anomaly.

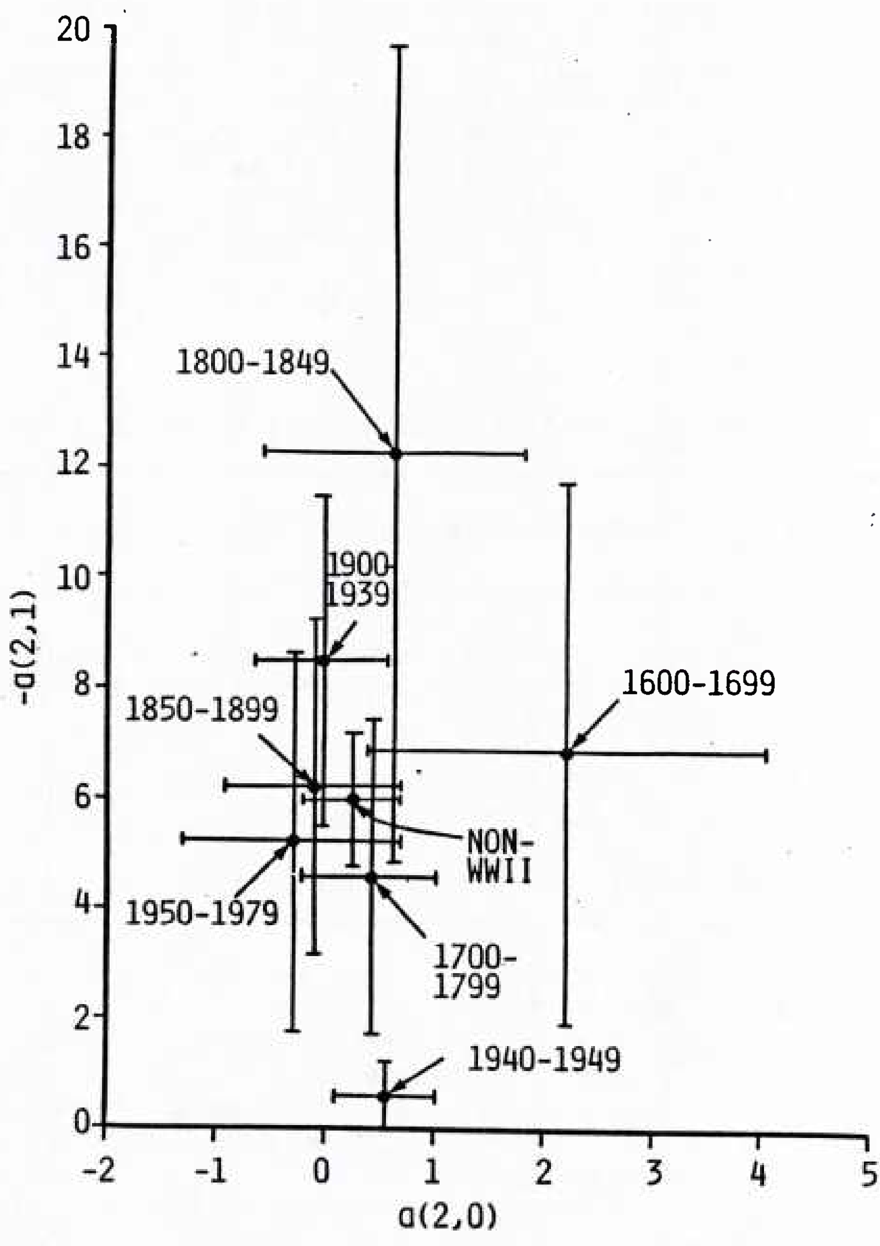


Figure 4-13. Means and Two-Standard Deviation Confidence Bands for the Logistic Regression Parameters a(2,0) and a(2,1)

d. Hypothetical Explanations of the World War II Anomaly

(1) Preliminary Remarks. Table 4-6 lists some possible explanations of the WWII anomaly to guide efforts to localize its source. In this writer's opinion, the first of these hypothesis--that the WWII data are flawed--is sufficiently more plausible than the others that it should receive by far the most effort over the near term, while work on the others should be held in abeyance pending the results of those efforts. This opinion was arrived at by a process of elimination, which is outlined below. In the first place, although Hypotheses 4 and 5 could perhaps be checked using data bases other than HERO's, such extensive use of other data bases was not within the scope of the effort reported in this paper. Moreover, neither Hypothesis 4 or 5 seems very plausible. It is difficult to see just how they could account for either the timing or the magnitude of the observed shifts in the logistic regression coefficient a(2,1). Accordingly, we direct our attention to Hypotheses 1 through 3.

Table 4-6. Hypothetical Explanations of the WWII Anomaly

1. The WWII data are flawed
2. The WWII data are correct, but their analysis is flawed
3. The WWII data and their analysis are correct--normal battle dynamics actually did change around 1940, but then changed back again before 1967
4. The WWII data and their analysis are correct, but the non-WWII data or their analysis is flawed
5. Both the WWII and the non-WWII data or analysis are flawed

(2) Comments on Hypothesis 2. Until some specific flaw in the analysis can be pinpointed, Hypothesis 2 remains purely ad hoc. Obviously, if there were any known flaws in either the theoretical analysis of Appendix J or in the DALOFIT computer program that reduces that theory to a practical computational scheme, they would already have been corrected. Besides, the hypothesis that there is a hidden flaw in the analysis--specifically one that causes the logistic regression parameter a(2,1) to shift back and forth at just the times and in the amounts observed--seems rather far fetched.

(3) Comments on Hypothesis 3. If battle dynamics actually did change around 1940, then it appears from Figures 4-11, 4-12 and 4-13 that it changed back again before the beginning of the Six Day Arab-Israeli War of 1967--all of the post-WWII battles in the HERO data base took place from 1967 to 1973 (see Appendix H). Table 4-4 also suggests that the relation between ADV and victory in battle has been stable for a very long time. Until some really excellent reasons are offered as to why the logistic regression coefficient a(2,1) should shift back and forth at just the times and in the amounts observed, Hypothesis 3 remains purely ad hoc. We shall also see in the next paragraph that the most anomalous battles are not distributed more or less evenly through the WWII subset, but that instead they tend to appear in clusters. However, this behavior is hard to explain on the basis of Hypothesis 3, and appears to require further ad hoc hypotheses to explain why the phenomenon turns on and off in the way the clustering of anomalous battles seems to indicate.

e. The Leading Hypothesis

(1) Preliminary Remarks. Based on the foregoing discussion, the currently most plausible hypothesis is that there are some flaws in the WWII data. Since there may also be flaws in some of the non-WWII data, a more precise statement of Hypothesis 1 is that the World War II data subset may have a noticeably higher percentage of battles with anomalous data than do the other data sets. Furthermore, experience has shown that when data are affected by errors, the anomalous data items often exhibit a "spotty" behavior, i.e., the anomalous data items tend to appear in clusters, so that certain subsets of the data have more than the average fraction of anomalous data items. Accordingly, we should select an indicator of anomalous data, see whether it occurs in the WWII data subset more frequently than in the others, and determine whether its occurrence tends to be spotty.

(2) An Indicator of Anomalous Data. The loglikelihood value of the outcome of an individual battle is the only indicator of anomalous data used within the scope of the effort reported in this paper. The choice of the loglikelihood value as an indicator of anomalous data has considerable statistical justification (see for example **Refs 4-1 and 4-2**, and many other standard statistical textbooks). By reference to Equation (J-13) of Appendix J, the value of this indicator for a particular battle is defined to be

$$L = LOG\ (P_{WINA}\ (ADV))$$

where

ADV is the observed defender's ADV value for the battle,

WINA is the observed outcome of the battle, i.e., WINA = +1, -1, or 0 according as to whether the attacker won, the defender won, or the outcome was a draw,

$P_r(x)$ is the probability that WINA = r for a battle in which ADV = x, where the probability is computed from some theoretical or fitted equation.

This may be expressed in words as follows. Calculate the theoretical or fitted probability of the occurrence of WINA, the observed outcome of the battle. L, the natural logarithm of this probability, is the loglikelihood value for that battle (with respect to the theoretical or fitted equations used to compute the probability of WINA). Some examples may help clarify the use of loglikelihood values as indicators of anomalous data.

Example 1 - Suppose that some theory predicts that the defender will invariably win whenever ADV is positive, and that we observe a battle in which the defender loses even though ADV is positive. Obviously, the hypothesized observation flatly contradicts the hypothesized theory. Here the observed outcome is DEFWIN, the probability of which is zero. Hence the loglikelihood value for this battle is

$$L = LOG\ (P_{WINA}\ (ADV)) = LOG\ (0) = -\ \text{infinity},$$

which corresponds to such an extremely anomalous observation, with respect to its theoretically predicted probability of occurrence, as to thoroughly discredit the theory.

Example 2 - Suppose that $P_r(x)$ is fitted to the post-WII data subset using logistic regression with ADV as the independent variable. The battle of Mount Hermon I (ISEQNO 593) is recorded in this data subset as having been won by the defender, an outcome which--on the basis of its ADV value and the fitted logistic regression function--has a probability of 0.223, so the loglikelihood value for Mount Hermon I outcome is

$$L = LOG\ (0.223) = -1.50.$$

Only 25 out of the 211 usable post-WWII era battles have a more negative loglikelihood, so Mount Hermon I is in the most anomalous 12 percent of the post-WWII battles with respect to the logistic regression fitted to the post-1940 era (1940-1979) subset.

Example 3 - The battle of Hushniya (ISEQNO 591) is recorded as having been won by the attacker, an outcome which--on the basis of Hushniya's ADV value and the logistic regression function fitted to the post-1940 era data subset--has a probability of 0.689, so the loglikelihood value for Hushniya is

$$L = LOG\ (0.689) = -0.373.$$

Since 123 out of the 211 usable post-1940 era battles have more negative loglikelihoods, Hushniya is in the least anomalous 42 percent of the post-1940 era data subset with respect to the logistic regression fitted to the post-1940 era (1940-1979) subset.

(3) Remark on the Treatment of Drawn Battles. Because drawn battles rarely occur (only about 5 percent of the HERO data base battles are drawn), their loglikelihood values tend to be much lower than those of other battles, even when they are not otherwise anomalous. Thus, when assessing anomalous battles, it is often appropriate to omit draws. Where convenient, results are provided when draws are either omitted or included.

(4) Anomalous Battles for the Pre-1940 and Post-1940 Eras Relative to the ALL-HERO Subset. This paragraph presents some findings on anomalous battles relative to the logistic regression fit of WINA versus adjusted ADV for the all-HERO data subset (counting draws as draws and not forcing symmetry). There are 16 battles in the HERO data base that lack sufficient data to compute ADV, leaving 585 usable battles in the all-HERO subset. Nine of these 585 battles have loglikelihoods less than -3.0 and are not draws. All nine of them are from the Okinawa Campaign of World War II. An additional eight battles have loglikelihoods of -3.0 to -2.0 and are not draws. Five of these eight battles are from the Italian Theater of World War II. The other three consist of one each from the Northwest European Theater of World War II, the Eastern Front of World War II, and the Golan Front of the Arab-Israeli 1973 October Campaign. Thus, these 17 battles with loglikelihoods less than -2.0 and not drawn are all from the 1940-1979 subset. Moreover, 16 of them are from the 1940-1949 (World War II) era. Tables 4-7 through 4-10 consistently indicate that there is a significantly higher proportion of anomalous battles in the 1940-1979 subset as compared to the 1600-1939 subset, whether the cutoff loglikelihood is taken as -1.0 or as -2.0, and whether draws are included in the tabulation or not. Accordingly, the all-HERO subset is heterogeneous and should be separated into at least a pre-1940 and a post-1940 era, each of which individually is likely to be much more nearly homogeneous than is the all-HERO subset. Results based on such a decomposition of the all-HERO subset will be presented in paragraph (6) below. First, however, the anomalous battles of the post-1940 or 1940-1979 era will be examined a little more closely relative to the all-HERO subset.

Table 4-7. First Table of Anomalous Battles for the Pre-1940 and Post-1940 Eras[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
1600-1939	0	374	374	0.0
1940-1979	17	194	211	8.1
Total	17	568	585	2.9

[a]Here an anomalous battle is one that is **not drawn**, and whose loglikelihood is **less than -2.0** relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 28.24 at 1 DOF. The probability of a greater chi-square value by chance is about 3×10^{-6}.

Table 4-8. Second Table of Anomalous Battles for the Pre-1940 and Post-1940 Eras[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
1600-1939	20	354	374	5.3
1940-1979	27	184	211	12.8
Total	47	538	585	8.0

[a]Here an anomalous battle is one, **drawn or not**, whose loglikelihood is **less than -2.0** relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 9.15 at 1 DOF. The probability of a greater chi-square value by chance is about $2x10^{-3}$.

Table 4-9. Third Table of Anomalous Battles for the Pre-1940 and Post-1940 Eras[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
1600-1939	22	352	374	5.9
1940-1979	41	170	211	19.4
Total	63	522	585	10.8

[a]Here an anomalous battle is one that is **not drawn**, and whose loglikelihood is **less than -1.0** relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 24.38 at 1 DOF. The probability of a greater chi-square value by chance is about $8x10^{-7}$.

Table 4-10. Fourth Table of Anomalous Battles for the Pre-1940 and Post-1940 Eras[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
1600-1939	42	332	374	11.2
1940-1979	51	160	211	24.2
Total	93	492	585	15.9

[a]Here an anomalous battle is one, **drawn or not**, whose loglikelihood is **less than -1.0** relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 15.94 at 1 DOF. The probability of a greater chi-square value by chance is about 6×10^{-5}.

(5) **Anomalous Battles for Theaters and Campaigns of the Post-1940 Era Relative to the All-HERO Subset.** This paragraph presents some findings on anomalous battles of the 1940-1979 subset relative to the logistic regression fit of WINA versus adjusted ADV for the all-HERO subset (counting draws as draws and not forcing symmetry). To obtain these results, the post-1940 era battles were grouped as indicated in Tables 4-11 and 4-12. This grouping was selected as a compromise between the following two principles:

(a) Each group's expected number of anomalous battles, estimated using the average frequency of anomalous battles in the post-1940 era, should be at least five. This is to make the application of the chi-squared test for independence in contingency tables more reliable. See, for example, pages 85 and 97 of **Ref 4-8.** (As there are too few battles with loglikelihood less than -2.0 to satisfy this principle, in Tables 4-11 and 4-12 anomalous battles are defined as those with loglikelihoods less than -1.0.)

(b) Each group of battles should be as homogeneous as possible. In practice, this means that they should be from the same theater and campaign, unless this seriously conflicts with principle (a) above.

Tables 4-11 and 4-12 show that in the post-1940 era the percentage of anomalous battles varies appreciably from one theater/campaign to another--in other words that the anomalous battles are "spotty" and tend to appear in clusters. This strongly suggests that errors may have crept into the data base for battles of the post-1940 era.

Table 4-11. First Table of Anomalous Battles for Theaters and Campaigns of the Post-1940 Era[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
North Africa, Misc., Tarawa, Iwo Jima	0	13	13	0.0
Italy (Salerno, Volturno)	8	21	29	27.6
Italy (Anzio, Rome, North Italy)	11	24	35	31.4
Northwest Europe	5	19	24	20.8
Eastern Front	3	26	29	10.3
Okinawa (7th Division)	5	12	17	29.4
Okinawa (96th Division)	4	7	11	36.4
1967 Six Day and 1968 Wars	0	20	20	0.0
1973 October War (Suez Front)	1	15	16	6.2
1973 October War (Golan Front)	4	13	17	23.5
Total	41	170	211	19.4

[a]Here an anomalous battle is one that is **not drawn**, and whose loglikelihood is **less than -1.0** relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 19.02 at 9 DOF. The probability of a greater chi-square value by chance is about 0.025.

Table 4-12. Second Table of Anomalous Battles for Theaters and Campaigns of the Post-1940 Era[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
North Africa, Misc., Tarawa, Iwo Jima	0	13	13	0.0
Italy (Salerno, Volturno)	10	19	29	34.5
Italy (Anzio, Rome, North Italy)	14	21	35	40.0
Northwest Europe	6	18	24	25.0
Eastern Front	4	25	29	13.8
Okinawa (7th Division)	5	12	17	29.4
Okinawa (96th Division)	4	7	11	36.4
1967 Six Day and 1968 Wars	2	18	20	10.0
1973 October War (Suez Front)	1	15	16	6.2
1973 October War (Golan Front)	5	12	17	29.4
Total	51	160	211	24.2

[a]Here an anomalous battle is one, drawn or not, whose loglikelihood is less than -1.0 relative to the logistic regression fit for WINA versus adjusted ADV using the all-HERO subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 18.72 at 9 DOF. The probability of a greater chi-square value by chance is about 0.028.

(6) **Anomalous Battles for Theaters and Campaigns of the Post-1940 Era Relative to the Pre-1940 Era.** This paragraph presents some findings on anomalous battles of the 1940-1979 subset relative to the logistic regression fit of WINA versus adjusted ADV for the 1600-1939 subset (counting draws as draws and not forcing symmetry). To obtain the first of these results, the post-1940 era battles were again grouped as indicated in Tables 4-11 and 4-12. The results are given in Tables 4-13 and 4-14. These tables show again that in the post-1940 era the percentage of anomalous battles varies appreciably from one theater/campaign to another--in other words that the anomalous battles are "spotty" and tend to appear in clusters. This is also visible in Figure 4-14. As before, this strongly suggests that errors may have crept into the data base for battles of the post-1940 era. To verify that the clustering of anomalous battles was not artificially induced by the specific groupings used in Tables 4-11 through 4-14, the run test was used **(see Refs 4-8 and 4-9)**. Two such tests were made. In both of them, battles of the post-1940 era were taken in the order in which they are listed in the HERO data base and in Appendix H, and all battles--including draws--were included. New runs were started each time the loglikelihood value crossed a preselected level. The first test used -1.0 as the preselected level, while the second test used -2.0 as the preselected level. In the first test, it was observed that 151 loglikelihood values were below -1.0 and 60 were above it, while 75 runs occurred--a value so low that a lower value would have occurred by chance only about 2 percent of the time. In the second test, it was observed that 165 loglikelihood values were below -2.0 and 46 were above it, while 61 runs occurred--a value so low that a lower value would have occurred by chance only about 1 percent of the time. As before, we conclude that the high and low loglikelihood values are "spotty" and clustered far more than would be at all likely by chance.

Table 4-13. Third Table of Anomalous Battles for Theaters and Campaigns of the Post-1940 Era[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
North Africa, Misc., Tarawa, Iwo Jima	0	13	13	0.0
Italy (Salerno, Volturno)	13	16	29	44.8
Italy (Anzio, Rome, North Italy)	11	24	35	31.4
Northwest Europe	6	18	24	25.0
Eastern Front	5	24	29	17.2
Okinawa (7th Division)	5	12	17	29.4
Okinawa (96th Division)	4	7	11	36.4
1967 Six Day and 1968 Wars	0	20	20	0.0
1973 October War (Suez Front)	1	15	16	6.2
1973 October War (Golan Front)	5	12	17	29.4
Total	50	161	211	23.7

[a]Here an anomalous battle is one that is not drawn, and whose loglikelihood is less than -1.0 relative to the logistic regression fit for WINA versus adjusted ADV using the 1600-1939 subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 23.54 at 9 DOF. The probability of a greater chi-square value by chance is about 0.005.

Table 4-14. Fourth Table of Anomalous Battles for Theaters and Campaigns of the Post-1940 Era[a]

Data subset	Number anomalous	Not anomalous	Total	Percent anomalous
North Africa, Misc., Tarawa, Iwo Jima	0	13	13	0.0
Italy (Salerno, Volturno)	15	14	29	51.7
Italy (Anzio, Rome, North Italy)	14	21	35	40.0
Northwest Europe	7	17	24	29.2
Eastern Front	6	23	29	20.7
Okinawa (7th Division)	5	12	17	29.4
Okinawa (96th Division)	4	7	11	36.4
1967 Six Day and 1968 Wars	2	18	20	10.0
1973 October War (Suez Front)	1	15	16	6.2
1973 October War (Golan Front)	6	11	17	35.3
Total	60	151	211	28.4

[a]Here an anomalous battle is one, drawn or not, whose loglikelihood is less than -1.0 relative to the logistic regression fit for WINA versus adjusted ADV using the 1600-1939 subset when draws are counted as draws and symmetry is not forced. This table's chi-square is 24.01 at 9 DOF. The probability of a greater chi-square value by chance is about 0.004.

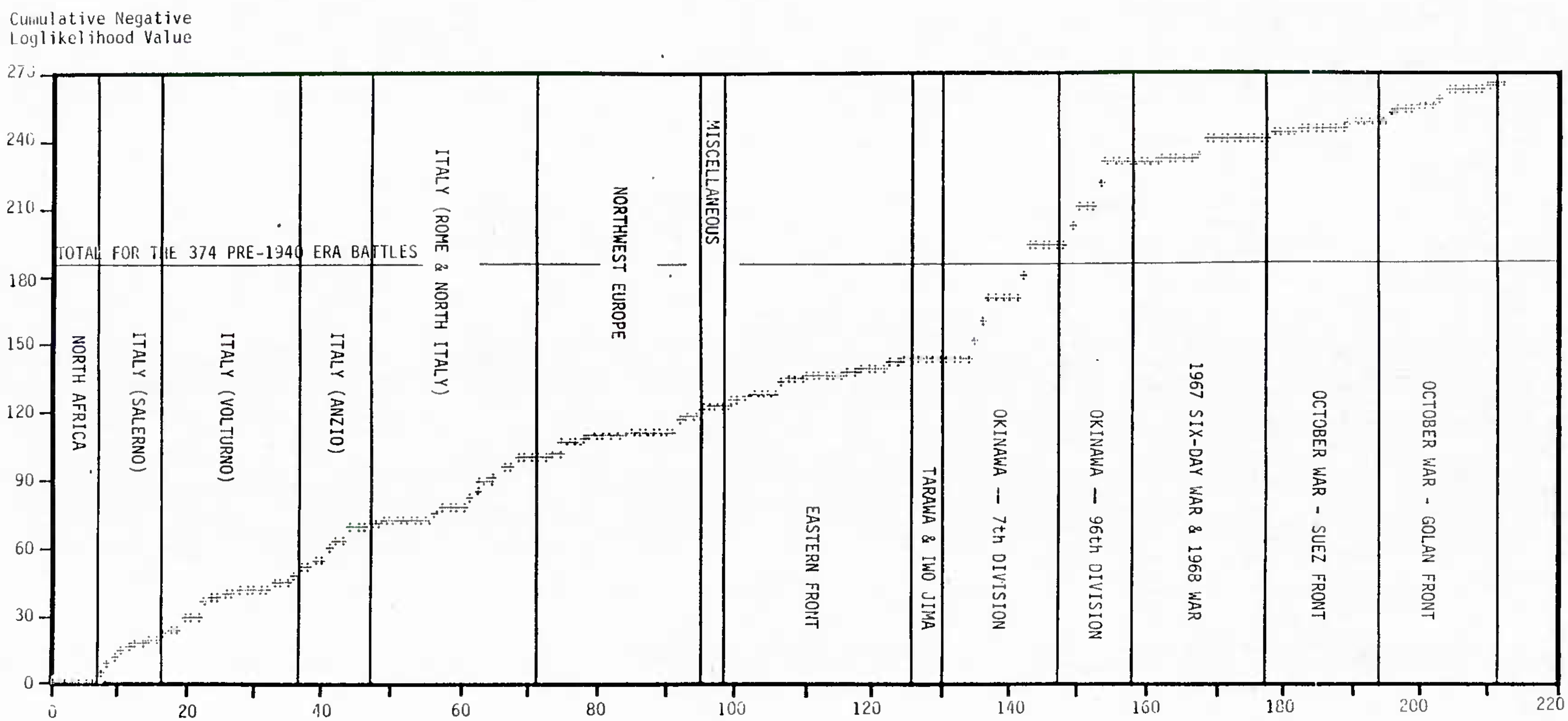


Figure 4-14. Cumulative Negative Loglikelihood for Post-1940 Era Battles Relative to the Logistic Regression Fit of WINA Versus ADV for the Pre-1940 Era, with Draws Counted as Draws and Symmetry Not Forced

(7) Other Attempts to Localize the Source of the World War II Anomaly. Some other attempts were made to localize the source of the World War II anomaly. It was reasoned that, if anomalous battles were due to substantial errors in their strength and loss data, this might be reflected in a tendency for anomalous battles to have an unusually high frequency of problem reports (see Chapter 2 for a discussion of problem reports). However, it was found that this was not the case--in fact anomalous battles tend to have fewer problem reports than nonanomalous battles. Apparently, whatever data flaws might be affecting the anomalous battles, this is not reflected in the problem reports. An analysis was also made of the sources used in the HERO data base for the Italian Campaign, to see whether some particular source was regularly associated with anomalous battles. However, since multiple sources were cited, it was not possible to tell just which sources were used for strengths and losses. Nor was it possible to find a single source that was consistently related to anomalous battles. Other attempts were made to isolate the source of the World War II anomaly by examining various subsets of the post-1940 era battles. However, the sample sizes were too small to reliably detect any statistical differences that might have been present.

(8) **Concluding Observations.** It has been shown that the post-1940 era battles differ significantly from the pre-1940 era battles with respect to the dependence of victory on ADV. This is called the World War II anomaly, since it starts with World War II and involves mainly WWII battles. However, most post-1940 era battles are not anomalous--relative to the logistic regression fitted to the pre-1940 era battles, about 72 percent of the post-1940 era battles have loglikelihoods above -1.0, and about 66 percent have loglikelihoods above -0.5 (see Figure 4-15).

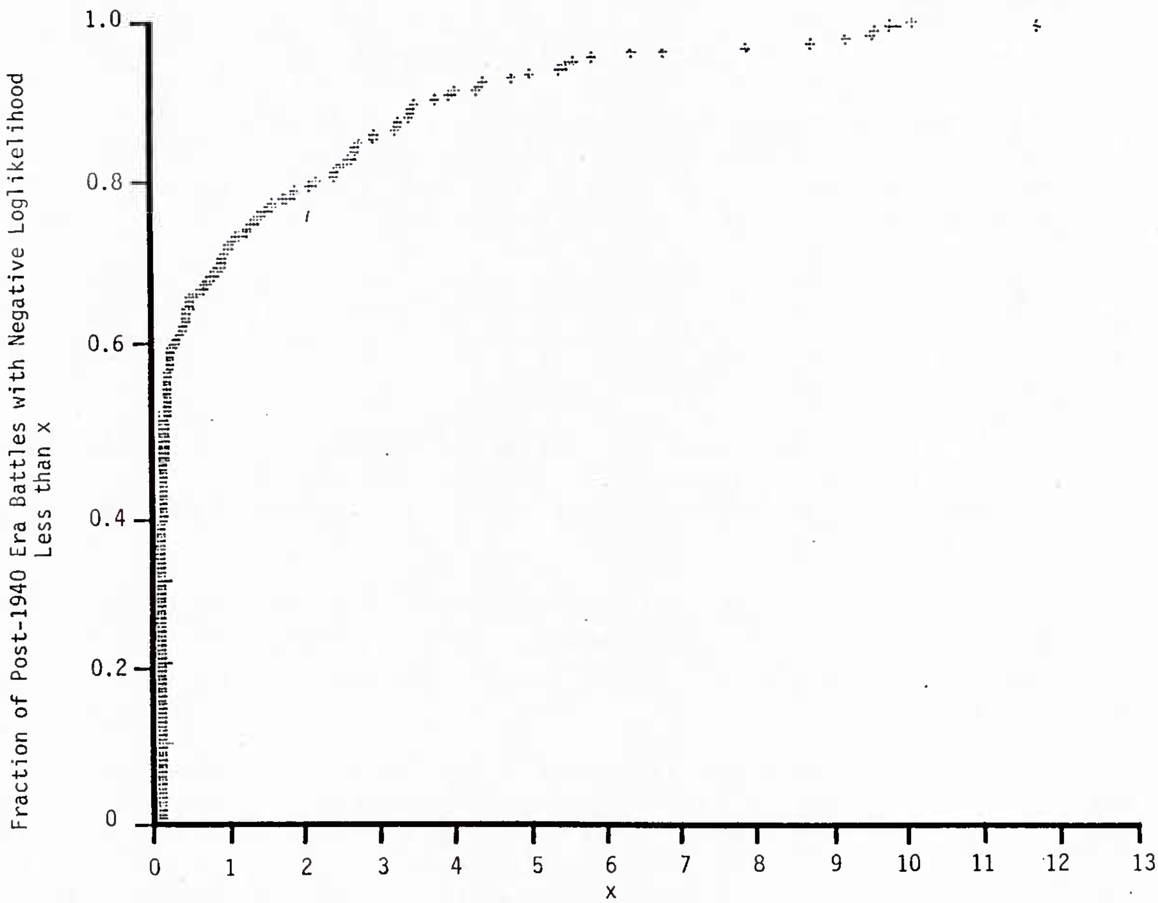


Figure 4-15. Distribution of Negative Loglikelihoods for Post-1940 Era Battles, Relative to the Logistic Progression Fit of WINA Versus Adjusted ADV for the Pre-1940 Era with Draws Counted as Draws and Symmetry Not Forced

Moreover, runs and contingency table analyses have shown that the anomalous battles in the post-1940 era occur spottily, rather than being spread more or less uniformly throughout the 1940-1979 data subset. For example, the Okinawa battles are highly anomalous but the battles of Tarawa and Iwo Jima are not--nor, according to Table 4-4, are some other Pacific Ocean island battles of World War II. The Italian Theater battles tend to be more anomalous than those of the Eastern Front. The 1967 Six Day and 1973 October War (Suez Front) battles are not particularly anomalous, but the 1973 War (Golan Front) battles are. Clearly additional effort will be needed to explain the peculiarities of the World War II anomaly.

4-5. NEXT STEPS

a. Next Steps for the World War II Anomaly

(1) Steps Currently Under Way. Several steps are currently under way to help resolve the status of the World War II anomaly, although their results were not available in time to be used in this paper. As mentioned in Chapter 2, the CDES Contract calls for HERO to, among other things:

(a) Clarify the total strength data. This will allow a sounder approach to judging when total strengths represent initial strengths and, when they do not, will help to indicate what procedure would be most effective in analyzing those strength data.

(b) Clarify the basis for assigning victory. This should help to clarify questionable assignments of victory, and can be used to determine whether victory in anomalous battles tends to be assigned on a different basis than for the other battles.

(c) Weight the strength and loss data according to the adjudged accuracy of the available information. This will help to indicate whether anomalous battles are usually among those with the less certain strength and loss data.

(d) Review selected strength, loss, and victory assessment values. HERO was asked to review carefully their assessments of strengths, losses, and victory for a list of selected battles. Although HERO was not told of the fact at the time, this list of battles was based on those found to be anomalous, i.e., as having unusually low loglikelihood values. This review will help to assure that the data provided by HERO for those battles is as accurate as HERO can make it.

(e) As of this writing, CAA plans to request proposals to conduct an independent review/reassessment of the strengths, losses, and victory values for anomalous HERO battles. This contract will help to determine the extent to which the data for these battles depends on the investigator.

(2) Other Steps. Several suggested next steps for the World War II anomaly are presented in Table 4-15. They all involve attempts to localize the source of the anomaly, and to understand its nature and causes.

Table 4-15. Next Steps For World War II Anomaly

1. Reassess the situation as CDES results become available
2. Try to localize the source of the anomaly, e.g.,
   a. Do the outliers tend to involve the same
      (1) Military units?
      (2) "Sources consulted"?
      (3) Locale, sector, campaign or theater?
      (4) Historical analyst?
      (5) Research agency?
   b. Do outliers tend to have more "problem reports" than other battles?
3. What happens if Italian and Okinawan campaign data are omitted?
4. Does the logistic regression fit converge as outliers are eliminated?
5. Interpretation and documentation of findings

b. Next Steps for Factors Associated with Victory. Table 4-16 presents some of the next steps that can be taken in the search for the factors associated with victory.

Table 4-16. Next Steps for Factors Associated with Victory

1. Redo the calculations as CDES results become available
2. What data subsets to use hinges on resolution of World War II anomaly
3. Do the findings extend to other data bases, e.g., BWS, CORG, air battles, sea battles, or wars?
4. Refine the choice of independent variables, e.g., ADV, RESADV, LOG (FER), and/or others
5. Refine choice of functional form for PWIN, e.g., logistic regression, probit regression, or others
6. Does the degree or decisiveness of victory become more pronounced at extreme ADV values?
7. Are ADV and EPS truly independent quantities?
8. Can ADV be predicted beforehand?
    a. Reduction of dimensionality
    b. Factor analysis
    c. Regression and correlation analysis
    d. Are casualties caused <u>after</u> defeat, or <u>before</u>?
    e. Does EPS have <u>any</u> influence on PWIN?
9. Can a simple linear weighting of infantry, artillery, tanks, and air be found that strongly influences PWIN?

First, the computations in this chapter need to be redone as the CDES contract results become available. Second, what data subsets to use hinges to some extent on the resolution of the World War II anomaly. Third, it is desirable to know whether findings based on an analysis of the HERO data base extend to other data bases. It may also be possible to refine the choice of independent variables--or of the functional form employed--in such a way as to improve the quality of the logistic regression results.

If ADV is a good measure of advantage in battle, then the degree or decisiveness of the victory should become more pronounced at extreme values of the ADV parameter. Theoretically, EPS and ADV should be independent of each other--to what extent is this borne out by the data? A key problem is to find some way of forecasting what the value of the ADV parameter will be before the battle starts, rather than relying on information about the losses taken during the battle. Tackling this problem will probably require the elimination of redundancy among the subjective variables listed in the HERO data base--some work along these lines has been started and is reported in Chapter 5. Finally, we might attempt to find some function of the force mix that can be used to predict the probability of winning, either directly or via its effect on losses--**Ref 4-10** describes a technique that may be useful for this purpose.

4-6. CONCLUDING OBSERVATIONS ON FACTORS ASSOCIATED WITH VICTORY

a. The variables ADV, LOG(FER), RESADV, LOG(CER), LOG(EPS), and LOG(FR) were compared with regard to the closeness of their association with victory in non-WWII battles, and were found to rank (from most closely associated to least) in the order listed. ADV, LOG(FER), and RESADV are nearly equally closely associated with victory in battle. The association between LOG(FR) and victory is not as close as any of the other five variables examined.

b. Some of the battles in the HERO data base are anomalous in the sense that their outcomes differ sharply from what is anticipated on the basis of the association of victory with ADV. A high proportion of the anomolous battles took place in the post-1940 era, even though most of those battles are not anomalous. In particular, the Italian, Northwest Europe, Okinawan, and 1973 October War (Golan Front) campaigns all seem to have relatively high incidences of anomalous battles. But the North Africa, Tarawa, Iwo Jima, Eastern Front, 1967 Six Day and 1968 Wars, and 1973 October War (Suez Front) campaigns all seem to have about the same proportion of anomalous battles as do the battles of the pre-WWII era. Various hypotheses as to the cause of these WWII anomalies were presented and discussed. While the issue has not been definitively resolved, internal and circumstantial evidence suggests that the WWII anomalies could well be due to flaws in the data, particularly for some of the post-1940 battles. Making an independent review and reassessment of the data on the anomalous battles would help greatly to provide data on which to base a determination of whether the WWII anomaly is a reflection of flawed data, or of some previously unanticipated phenomenon.

c. Despite the WWII anomaly issue, ADV (or, alternatively, LOG(FER)) has been shown both theoretically and empirically to be substantially more accurate than other figures of merit for comparing the "military worth" of alternative materiel, organizations, and tactics.

CHAPTER 5

ANALYSIS OF REDUNDANCY

5-1. INTRODUCTION. The HERO data set provides 601 land battles for statistical analysis. With each battle, there are 29 variables that we will focus on. Many of these variables are judgmental in nature. The correlations among these variables are high, thus making regression and some other statistical analyses difficult. Some of the variables include or at least partially overlap others, thus wholly or partly duplicating information. For example, in HERO's Table 4, combat effectiveness is defined as "a complex factor, subsuming--among other elements--leadership, training and experience, morale, and logistics." Hence, CEA at least partially overlaps LEADA, TRNGA, MORALA, LOGSA, and unspecified other variables. Accordingly, we expect CEA to be correlated statistically with these other variables, and consequently to be at least partly redundant. Similarly, MORALA and LEADA may be correlated since capable leadership is conducive to high morale and inept leadership to poor morale. These are instances of duplication of information, which is not an unusual situation in complex data sets. Along with the so-called "curse of high dimensionality" comes the problem of "redundancy in information." To cope with them requires a method for reducing the dimensionality of the data base without sacrificing information contained in it. The notions of dimensionality and information in the data base need explanation. The term dimensionality of data base refers here to the number of variables, 29 in the present case. We shall show how these 29 variables, or observables in the standard terminology, can be expressed as a linear combination of 8 underlying variables or factors. The nature of the 29 observables may be characterized collectively by the variances of the observables and the correlations between each pair of the observables. In other words, a 29 * 601 matrix--29 observed values for each battle--can be summarized by 29 variances of each observable and the table of correlations of each variable with the others, which has 29 * (29-1)/2 = 29 * 14 entries. Either the correlation table or the corresponding table of covariances can be used. In this chapter, the contents of the variance-covariance matrix is called the information in the data base. It will be shown that eight factors can be so chosen that they (i) account for all the correlations; (ii) among all the possible linear combinations of the observables, the same eight factors account for the maximum of the sum of the variances of the observables; (iii) moreover, the eight factors are uncorrelated among themselves (which is an important consideration for subsequent statistical analysis work). The method chosen for this purpose is factor analysis **(Refs 5-1, 5-2, 5-3, 5-4)**.

5-2. FACTOR ANALYSIS. The statistical technique of factor analysis was used for this dimension reduction. The 29 variables chosen for the application of factor analysis are listed in Table 5-1, along with their means and standard deviations for the exploratory subsample (as indicated by the sample size). The classical technique known as "principal factoring" **(Ref 5-1)** will be applied to these 29 variables.

Table 5-1. Variables Selected for Factor Analysis

Variable	Mean	Standard dev	Sample size
SURPA	.4200	.8053	100
CEA	.1400	.5854	100
TRNGA	-.0800	.5446	100
MORALA	.2400	.5527	100
LOGSA	-.0100	.4143	100
LEADA	.2300	.8860	100
SURPAA	.2700	.5478	100
AEROA	.1500	.4578	100
INITA	.6500	.5573	100
WINA	.3200	.9307	100
KPDA	1.7702	3.4811	94[a]
QUALA	.1400	.5508	100
ACHA	6.2100	2.3540	100
MOMNTA	.1400	.3437	100
INTELA	.1200	.4774	100
TECHA	.0100	.1738	100
ACHD	5.0306	2.1174	98[a]
RESA	.0600	.6000	100
MOBILA	.1300	.3667	100
AIRA	.1000	.3892	100
FPREPA	.2000	.5505	100
WXA	-.0300	.3320	100
TERRA	-.3700	.5624	100
LEADAA	.2200	1.0404	100
PLANA	.3000	.6590	100
MANA	.0900	.3786	100
LOGSAA	.0300	.2227	100
FORTSA	-.4700	.5588	100
DEEPA	-.2000	.4020	100

[a]Some of the battles in the exploratory subsample are missing these data items.

a. Classical Factor Analysis. Under this model each of the (observable) variables is assumed to be a linear function of a small number of hypothesized common factors and a single unique factor **(Ref 5-2)**. Under this model, the common factors generate the correlations observed among the original variables, while the unique term contributes only to the variance of the particular variable. In mathematical symbols, we have the following expression **(Ref 5-3)**:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + .. + a_{jm}F_m + U_j. \tag{5-1}$$

In Equation (5-1), $j = 1, 2, \ldots, n$ where n is the number of (observable) variables. For our application, $n = 29$. Also, m is the number of (unobservable) common factors. Z_j is the value of the (observable) variable j in standardized form, i.e., zero mean and unit standard deviation. For $i = 1, 2, \ldots, m$ F_i is the ith (unobservable) common factor introduced to account for the correlations among the Z_j. U_j is the unique factor introduced to account for the variance of Z_j. And the a_{ji} are the standardized multiple regression coefficients of variable j on factor i (factor loadings).

The following conditions are imposed on the hypothesized factors **(Ref 5-1)**:

$\text{Corr}(F_i,F_j) = 0$ for $i \neq j$ and $i, j = 1, 2, \ldots, m$.

$\text{Corr}(F_i,U_j) = 0$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, 29$.

$\text{Corr}(U_i,U_j) = 0$ for $i \neq j$ and $i,j = 1, 2, \ldots, 29$.

Since the common factors $F_1, F_2, \ldots, F_m$ are uncorrelated, it follows from Equation (5-1) that

$$\text{Corr}(Z_j,Z_k) = \sum_{i=1}^{m} a_{ji}a_{ki} \text{ for } j \neq k \text{ and } j,k = 1, 2, \ldots, 29.$$

b. Number of Factors. Under the classical factor model, one must decide how many factors m to postulate to account for the correlations among the set of original variables Z_j. We shall choose the procedure based on the eigenvalues of the correlation matrix of the original variables Z_j. The rule is: select as many factors m as there are eigenvalues greater than one. The rationale for this procedure is as given in the next paragraph.

c. Eigenvalues and Factors. The number m of factors F_i in Equation (5-1) can be determined by the magnitudes of the eigenvalues of the correlation matrix. Define the total variance in the data to be the sum of the variances of each variable. Since each standardized variable Z_j has unit variance, the total variance is equal to the number of variables, or 29 in this case. Let the eigenvalue associated with factor i be V_i where $i = 1, 2, \ldots, m$. In the principal-component method, V_i can be shown to be the contribution of factor i to the total variance. On the average, this proportion is equal to m/29. Factors are rated in importance by the ratio

$V_i/29$. Only those factors are retained in Equation (5-1) whose associated eigenvalues are at least equal to 1. Thus, only those factors are retained which account for at least as much variance as any single variable in the data. For details, see **Ref 5-1**.

5-3. THE DATA SET AND EXPLORATORY SUBSAMPLE. As mentioned above, the data consists of 601 battles, and each battle is described by 80 or 90 data items. See Appendices E, F, G, and the Glossary for a description of these data items. From the 601 battles, an exploratory subsample of 100 battles has been drawn for exploratory analysis. The battles in the exploratory subsample span the years from 1631 to 1942 A.D. The exploratory subsample thus covers a broad range of battles representative of the pre-World War II HERO data base. We shall use the exploratory subsample to estimate factor loadings and factor score coefficients. These estimates of a_{ji} obtained from the exploratory subsample are also useful in cross-validation. We will use these estimates to predict the variances and correlations of the remaining set of 501 battles and their 29 associated variables.

5-4. FACTORS FOR THE EXPLORATORY SUBSAMPLE

a. Approach. Initially, the 29 variables listed in Table 5-1 were analyzed using the exploratory subsample. A part of their correlation table is shown in Table 5-2. Table 5-3 shows the eigenvalues of the correlation matrix. We see that eight of the eigenvalues are greater than 1. Therefore, we shall postulate eight factors in Equation (5-1). These eight factors account for 70.9 percent of the total variance (sum of 29 variances). The rest of the factors are not significant contributors to the total variance. Factor 9, for example, contributes only 3.3 percent to the variance of any of the variables. The rest of the factors contribute even less. The eight factors postulated also account for the correlation among the 29 variables. We have reduced the 29 variables to eight factors without significantly losing any information in the exploratory subsample. That is, the eight factors account for much of the variances and correlations of the exploratory subsample data. We also note from Table 5-3 that factor 1 accounts for 24.0 percent of the total variance, Factor 2 accounts for 14.5 percent. In practice, the factors are ranked in importance according to the amount of variance accounted for by them, that is, the factors are ranked according to their corresponding eigenvalues; in analysis and graphical representations, the most important factor is examined first, then the next, and so on.

Table 5-2. Correlation Between Variables for the Exploratory Subsample, n = 100

	SURPA	CEA	TRNGA	MORALA	LOGSA
SURPA	1.00000	.30167	.30734	-.02448	.22439
CEA	.30167	1.00000	.60480	-.10473	.17215
TRNGA	.30734	.60480	1.00000	-.47253	.26506
MORALA	-.02448	-.10473	-.47253	1.00000	.05471
LOGSA	.22439	.17215	.26506	.05471	1.00000
LEADA	.30175	.52069	.41536	-.11387	.19898
SURPAA	.88410	.16414	.24243	-.01601	.19006
AEROA	-.06294	-.11664	-.15396	.41518	.11451
INITA	.28548	.27509	.10650	.20988	.20344
WINA	.26328	.30575	.15066	.22229	.29656
KPDA	.18373	.29743	.05339	.22142	.04579
QUALA	.34387	.65796	.54280	-.04512	.09473
ACHA	.30431	.37365	.32054	.05404	.28184
MOMNTA	-.06754	.10077	.11276	.19076	-.06013
INTELA	.52382	.15589	.19272	-.03369	.21044
TECHA	-.03028	.18437	.00854	.29026	.14171
ACHD	-.43927	-.40648	-.21027	-.19863	-.18578
RESA	.17706	-.02412	-.04699	.38259	.44946
MOBILA	.25762	.14940	-.04856	.44262	.00864
AIRA	-.03862	-.15047	-.15248	.45075	.19419
FPREPA	.05917	-.24408	-.24934	.17264	.00886
WXA	.08529	.07369	.04246	-.01542	-.07565
TERRA	.21252	.12804	.13325	-.03640	-.05940
LEADAA	.38243	.47884	.29880	-.00492	.14577
PLANA	.36877	.28229	.15198	.13311	.15909
MANA	.40436	.07917	-.01372	.08883	.07020
LOGSAA	.26664	-.03249	.10328	-.05909	.55072
FORTSA	-.07309	-.01295	.10755	-.31793	-.02051
DEEPA	-.01870	.11998	.24915	-.50918	-.07278

	KPDA	QUALA	ACHA	MOMNTA	INTELA
SURPA	.18373	.34387	.30431	-.06754	.52382
CEA	.29743	.65796	.37365	.10077	.15589
TRNGA	.05339	.54280	.32054	.11276	.19272
MORALA	.22142	-.04512	.05404	.19076	-.03369
LOGSA	.04579	.09473	.28184	-.06013	.21044
LEADA	.30226	.34730	.65951	-.00719	.24456
SURPAA	.19504	.27515	.31590	-.04124	.49287
AEROA	.12958	-.12417	-.04627	.12021	-.08319
INITA	.27049	.25993	.58784	.15072	.31133
WINA	.41612	.36488	.80349	.14066	.29919

Table 5-3. Eigenvalues of the Correlation Matrix for the Exploratory Subsample, n = 100

Factor	Eigenvalue	Percent of variation	Cumulative percent
1	6.95060	24.0	24.0
2	3.69587	12.7	36.7
3	2.24469	7.7	44.5
4	1.86942	6.4	50.9
5	1.83626	6.3	57.2
6	1.42669	4.9	62.2
7	1.31924	4.5	66.7
8	1.22681	4.2	70.9
9	.95550	3.3	74.2
10	.86036	3.0	77.2
11	.76894	2.7	79.8
12	.67302	2.3	82.2
13	.56296	1.9	84.1
14	.53249	1.8	85.9
15	.49030	1.7	87.6
16	.45518	1.6	89.2
17	.43961	1.5	90.7
18	.43096	1.5	92.2
19	.41278	1.4	93.6
20	.29186	1.0	94.6
21	.26675	.9	95.6
22	.25700	.9	96.4
23	.22931	.8	97.2
24	.21432	.7	98.0
25	.16699	.6	98.5
26	.13967	.5	99.0
27	.10614	.4	99.4
28	.10027	.3	99.7
29	.07602	.3	100.0

b. Findings. Table 5-4 shows the varimax factor loadings for the eight factors for the subsample data. The factor loadings a_{ji} of Equation (5-1) are given for each factor and every variable. We shall use the factor loadings to come up with names for the factors, to give the factors an intuitive feel. The eight factors are influenced by those variables which have high values of a_{ji} (high loadings).

CMDA. Factor 1 loads heavily on LEADA, INITA, WINA, KPDA, ACHA, RESA, MOBILA, LEADAA and PLANA (for an explanation of these names, see the Glossary). These attributes are associated with the command structure of the attacker. Therefore, Factor 1 can be named as "Command Favoring the Attacker."

WINGSA. Factor 2 loads heavily on the variables MORALA, AEROA, and AIRA. These variables seem related to air power, so Factor 2 will be called "Wings Favoring the Attacker."

SHOCKA. This is the name assigned to Factor 3, since SURPA, SURPAA, INTELA, and MANA, which measure surprise and maneuver achieved by the attacker, are strongly dependent on Factor 3.

TRAINA for Factor 4, since CEA, TRNGA, LEADA, QUALA, and FPREPA all have high factor loadings on this factor.

SUPTA. Factor 5 expresses logistical support favoring the attacker, since variables LOGSA and LOGSAA load heavily on this factor.

SPEEDA for Factor 6; since KPDA, MOMNTA, and TERRA load heavily on this factor, we associate it with the attacker's speed and the pressure this puts on the defender.

AGILA variables WXA, MOBILA and MANA are prominent components of Factor 7 as measured by their factor loadings; these quantities measure the agility of the attacker.

EQUIPA. The variable TECHA is the dominant contributor to Factor 8, hence the name.

Table 5-4. Factor Loading for the Exploratory Subsample, n = 100 (varimax rotated factor matrix)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
SURPA	.24115	.06745	.85297	.24197	.08710	-.03742	.06764	-.09266
CEA	.39645	-03642	.07809	.72194	-.08382	.06649	.10445	.21378
TRNGA	.13811	-.15997	.11233	.83193	.17629	.10102	-.05497	-.06007
MORALA	.26130	.58603	-.00724	-.46780	-.08382	.06461	.13806	.32889
LOGSA	.19563	.13646	.07631	.15108	.79058	-.08053	-.14746	.09072
LEADA	.72108	-.07136	.04819	.40484	.04375	-.19021	-.07890	-.13428
SURPAA	.26311	.05957	.84320	.12914	.07136	-.03243	.05721	-.21608
AEROA	-.02150	.85158	-.11294	-.01805	.04911	.16180	-.19315	-.11396
INITA	.66923	.08081	.12332	.00517	.18594	-.06193	.14627	.23331
WINA	.83187	-.05812	.09694	-.06715	.22508	.16622	.01480	.03244
KPDA	.60274	.18501	.01577	.05794	-.20880	.38216	.01629	-.00463
QUALA	.38292	-.05657	.20292	.56517	-.13053	.24622	.06730	.24461
ACHA	.82633	-.08021	.06331	.15654	.23368	.09487	-.05171	-.04914
MOMNTA	.11903	.09131	-.00132	.00418	-.07521	.72780	.03564	.13799
INTELA	.19432	-.14149	.68227	.02767	.19945	.14182	-.26329	.09749
IECHA	.07979	.16753	-.07252	.06034	.03298	.16583	-.18262	.78047
ACHD	-.83622	.03878	-.27497	-.06889	.04608	.05372	-.00203	-.02094
RESA	.52245	.34788	-.00006	-.22936	.48158	.01349	.11347	-.11135
MOBILA	.37584	.22580	.25177	-.17884	-.01611	.10081	.49643	.03914
AIRA	.01548	.84312	-.16975	-.07758	.17989	.17027	-.23452	-.13426
FPREPA	.18133	.12977	.09810	.46145	-.03468	.20586	-.04369	-.52114
WXA	-.03482	-.05744	-.06804	.13055	.07554	.13284	.82799	-.17629
TERRA	-.05576	-.04451	.25354	.14130	.08927	.72375	.16001	.08073
LEADAA	.78601	-.06128	.14764	.28963	-.07546	-.14080	-.00341	-.09977
PLANA	.68601	.01053	.23972	.02837	.04224	.19819	.05390	.05458
MANA	.13689	-.09821	.55469	-.20115	.07037	.18106	.36691	.16485
LOGSAA	.03640	-.00716	.22297	-.02274	.82691	.03783	.19189	-.02437
FORTSA	.07754	-.62135	-.11760	-.04677	.11953	.21288	-.22761	-.24115
DEEPA	.06896	-.71711	-.06878	.15748	-.10157	.07689	-.20623	-.13738

5-5. VERIFICATION FOR THE EXPLORATORY SUBSAMPLE

a. Approach. The eight factors which account for most of the variance over all the variables are used to approximate these 29 variables. Equation (5-1) expresses the variables in terms of the factors. We now employ the "principal" component model, in which Equation (5-1) is simplified to

$$FZ_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{j8}F_8 \qquad (5\text{-}2)$$

where

FZ_j is the fitted value of the jth variable in its standardized form,

F_i is the ith factor,

a_{ji} is the ith factor loading for the jth variable (see Table 5-4 for values)

Replacement of 29 variables by eight factors is a considerable savings in data space; Equation (5-2) gives an insight into the mutual relationships of the variables.

b. Correlation. Since we have postulated a linear model, as expressed by Equation (5-2), it is possible to measure the goodness of fit between the approximation and the observed values by Pearson's correlation coefficient r. Table 5-5 gives the list of the variables and their correlations with their corresponding approximations (FZ_j from Equation (5-2), where the a_{ji} are taken from Table 5-4).

Table 5-5. Correlations Between Variables and Their Fitted Values for the Exploratory Subsample

For correlation between	Value of the correlation between Z_j and FZ_j
Z1(SURPA) and FZ1	.93
Z2(CEA) and FZ2	.87
Z3(TRNGA) and FZ3	.88
Z4(MORALA) and FZ4	.88
Z5(LOGSA) and FZ5	.86
Z6(LEADA) and FZ6	.87
Z7(SURPAA) and FZ7	.90
Z8(AEROA) and FZ8	.91
Z9(INITA) and FZ9	.69
Z10(WINA) and FZ10	.88
Z11(KPDA) and FZ11	.77
Z12(QUALA) and FZ12	.81
Z13(ACHA) and FZ13	.88
Z14(MOMNTA) and FZ14	.77
Z15(INTELA) and FZ15	.79
Z16(TECHA) and FZ16	.86
Z17(ACHD) and FZ17	.89
Z18(RESA) and FZ18	.84
Z19(MOBILA) and FZ19	.75
Z20(AIRA) and FZ20	.93
Z21(FPREPA) and FZ21	.76
Z22(WXA) and FZ22	.88
Z23(TERRA) and FZ23	.80
Z24(LEADAA) and FZ24	.70
Z25(PLANA) and FZ25	.73
Z26(MANA) and FZ26	.77
Z27(LOGSAA) and FZ27	.87
Z28(FORTSA) and FZ28	.77
Z29(DEEPA) and FZ29	.79

All the sample values of correlation are significantly different from zero. A statistical test of the hypothesis that the correlation is really zero is given by the statistic:

$$t = r((n-2)/(1-r^2))^{\frac{1}{2}}, \ \mathrm{DOF} = n-2.$$

For details, see **Ref 5-3.** From Table 5-5, we see that the minimum value of r is 0.70. The corresponding t-value is

$$t = 9.70 \text{ with } 98 \text{ DOF}$$

This value is significant at the one percent level. We conclude that the correlation is significantly different from zero. Since this statement is true for the minimum value of the sample correlations, it is true for all of the correlations in the above list. We also observe that all the correlations are close to 1, indicating a high correlation and therefore a close fit between the observed and estimated variables. In fact, we can test the hypothesis that they come from a population with a specific correlation other than zero. For such a test, use the Fisher's z-transform:

$$z = \tfrac{1}{2} \ln((1+r)/(1-r)).$$

Fisher's z has a normal distribution with variance equal to $1/(n-3)$ (see **Ref 5-5**). The smallest of the sample correlations is 0.70. Applying the z-transform, it is found that it could have come from a population with correlation 0.8. We conclude that the agreement between the observables and their estimates is very close.

c. Linear Fit. To check the linear fit between the observed and fitted variables for the exploratory subsample data, the equation

$$y = a+bz \tag{5-3}$$

was fitted to the data, where

y = one of the variables Z_j

z = fitted values FZ_j from Equation (5-2)

If the fitted line (Equation 5-3) is to express a proper relationship between y and z, then we should prove that the regression coefficient b is nonzero. This hypothesis is tested by the F-test **(Ref 5-6)**. The results for tests of the regression coefficients are given in Table 5-6. Note that all the F-statistics are greater than the critical value 6.90 at 1 percent level of significance. We conclude that z is a good predictor of y, and therefore the 29 observed variables can be adequately replaced by eight factors.

Table 5-6. Tests of Regression Coefficients for the Exploratory Subsample

Dependent variable (y in Eqn 5-3)	F(1,90)
SURPA	564.4
CEA	278.2
TRNGA	323.9
MORALA	321.7
LOGSA	265.1
LEADA	277.7
AEROA	407.4
SURPAA	394.0
INITA	81.0
WINA	320.0
KPDA	128.2
QUALA	169.2
ACHA	296.7
MOMNTA	127.1
INTELA	151.7
TECHA	263.3
ACHD	355.8
RESA	221.7
MOBILA	117.1
AIRA	603.4
FPREPA	120.6
WXA	310.6
TERRA	164.6
LEADAA	84.4
PLANA	102.3
MANA	130.0
LOGSAA	269.7
FORTSA	129.0
DEEPA	153.6

5-6. CROSS-VALIDATION. We have shown that the 29 variables can be replaced by eight factors for the exploratory subsample. The method can be extended to the rest of the data with 501 battles and the 29 variables. This procedure will also be useful in cross-validation of the data reduction procedure by the factor analytic method. We need the values of F_1, F_2, ... , F_8 for the 501 battles. We use the factor score coefficients given in Table 5-7 to calculate factor scores F_1, F_2, ... , F_8, then use them in Equation 5-2 to approximate the observed variables Z_j by their fitted values FZ_j. As explained in paragraph 5-5b, the verification of the fit is carried out by calculating the correlation coefficient r between the observed variables and their fitted values. These correlation coefficients for the nonexploratory subsample are given in Table 5-8.

Table 5-7. Factor Score Coefficients for the Exploratory Subsample, n = 100

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
SURPA	-.06136	.07771	.30054	.07685	-.05164	-.07434	-.02225	-.06083
CEA	.02600	.05750	-.03970	.29444	-.06103	-.00223	.08911	.08411
TRNGA	-.05403	.04190	-.02433	.36617	.09498	.06351	-.01072	-.09684
MORALA	.07502	.12348	.00978	-.18914	-.08138	-.00553	.05609	.21591
LOGSA	-.01559	.01702	-.04368	.05963	.42843	-.04287	-.08784	.07439
LEADA	.14875	.02328	-.07318	.13479	-.02406	-.14754	-.03378	-.12647
SURPAA	-.04291	.06946	.35635	.03003	-.06557	-.06596	-.03453	-.14039
AEROA	-.02665	.32301	-.02632	.11643	-.01449	.08670	-.13559	-.16065
INITA	.13225	-.01732	-.03344	-.05878	.06402	-.09426	.08420	.16350
WINA	.17677	-.08241	-.07234	-.11613	.08526	.06779	-.02325	.02844
KPDA	.12481	.05377	-.05468	.00113	-.15433	.19696	-.02631	-.05423
QUALA	.01429	.02632	.03077	.21364	-.09267	.10996	.03158	.11726
ACHA	.16713	-.05052	-.09441	-.00474	.08959	.03075	-.04844	-.04600
MOMNTA	-.00038	-.01249	-.07074	-.00857	-.02143	.44580	-.02156	.05806
INTELA	-.04151	-.06579	.31114	-.06010	.05131	.86116	-.26835	.10736
TECHA	-.00232	-.00857	-.00550	-.03917	.04396	.08250	-.13243	.53415
ACHD	-.17196	.02359	-.04303	.05512	.09557	.08754	.03081	-.01023
RESA	.10654	.06317	-.09965	-.06580	.22505	-.02247	.06337	-.08019
MOBILA	.05742	.04654	.04662	-.07701	-.85631	-.00974	.29359	.02143
AIRA	-.01940	.30384	-.03583	.08215	.85768	.09821	-.16652	-.16467
FPREPA	.06607	.01989	.03331	-.18355	-.06343	.13569	-.07844	-.35251
WXA	-.02983	-.00344	-.13578	.11920	.00434	.04148	.57561	-.14156
TERRA	-.09361	-.02231	.06534	.06337	.05224	.44715	.04589	-.07841
LEADAA	.16463	.01318	-.02257	.07024	-.10279	-.13247	-.00377	-.09631
PLANA	.12267	-.02639	.02372	-.05624	-.02747	.07701	-.00813	.02917
MANA	-.02977	-.08453	.21884	-.14673	.00211	.06348	.17431	.15828
LOGSAA	-.06249	-.05221	.00516	-.01202	.45569	.02428	.11807	.02283
FORTSA	.04728	-.24698	-.08704	-.10153	.10692	.19306	-.15780	-.12248
DEEPA	.04311	-.24307	-.04649	-.02496	-.01753	.09676	-.13103	-.05966

Table 5-8. Correlations Between Variables and Their Fitted Values for the Nonexploratory Subsample

For correlation between	Value of the correlation coefficient between Z_j and FZ_j
Z1(SURPA) and FZ1	.54
Z2(CEA) and FZ2	.59
Z3(TRNGA) and FZ3	.48
Z4(MORALA) and FZ4	.23
Z5(LOGSA) and FZ5	.42
Z6(LEADA) and FZ6	.05*
Z7(SURPAA) and FZ7	.34
Z8(AEROA) and FZ8	.45
Z9(INITA) and FZ9	.38
Z10(WINA) and FZ10	.46
Z11(KPDA) and FZ11	.49
Z12(QUALA) and FZ12	.60
Z13(ACHA) and FZ13	.57
Z14(MOMNTA) and FZ14	.27
Z15(INTELA) and FZ15	.15
Z16(TECHA) and FZ16	.21
Z17(ACHD) and FZ17	.31
Z18(RESA) and FZ18	.36
Z19(MOBILA) and FZ19	.42
Z20(AIRA) and FZ20	.35
Z21(FPREPA) and FZ21	.07*
Z22(WXA) and FZ22	.16
Z23(TERRA) and FZ23	.30
Z24(LEADAA) and Z24	.10*
Z25(PLANA) and Z25	.60
Z26(MANA) and Z26	.58
Z27(LOGSAA) and Z27	.72
Z28(FORTSA) and Z28	.74
Z29(DEEPA) and Z29	.59

Variables marked with * show practically zero correlation with their fitted counterparts, the other variables are correlated with their fitted variables FZ_j. The test of significance has been carried out using Fisher's z-test on correlation r as explained in paragraph 5-5b. It should be observed that the values of r are lower in this list than the corresponding values for the exploratory subsample. This is to be expected, as the factor loadings and factor score coefficients for Table 5-8 were estimated from the exploratory subsample data. The "goodness" of fit was measured by the F-statistics which are given in Table 5-9 for the nonexploratory subsample of the data.

Table 5-9. Tests of Regression Coefficients for the Nonexploratory Subsample

Dependent variable (z in Eqn 5-3)	F(1,447)
SURPA	196.5
CEA	250.0
TRNGA	141.4
MORALA	27.8
LOGSA	100.9
LEADA	1.2*
AEROA	122.8
SURPAA	62.3
INITA	79.7
WINA	126.9
KPDA	153.2
QUALA	267.8
ACHA	233.2
MOMNTA	37.7
INTELA	11.5
TECHA	21.5
ACHD	49.6
RESA	71.6
MOBILA	102.1
AIRA	68.3
FPREPA	2.6*
WXA	12.9
TERRA	45.8
LEADAA	4.8*
PLANA	266.0
MANA	238.5
LOGSAA	535.0
FORTSA	570.8
DEEPA	258.6

Three variables--LEADA, FPREPA, and LEADAA--show poor F-value (critical F-value is 6.7 at 95 percent confidence level). The rest of the 29 variables exceed the critical value and therefore the approximation of these variables by eight factors can be regarded as adequate. This completed the cross-validation of the technique of factor analysis for data reduction.

5-7. CONCLUSION. We have shown that it is feasible to replace 29 variables with eight factors. Moreover, every observed variable can be generated by the use of approximation formula (Eqn 5-2). The approximation is linear, therefore simple to comprehend. Other advantages in this procedure are:

a. **Parsimony.** Instead of 29 variables, we have eight factors whose meanings are intuitively explainable.

b. **Orthogonality.** The observed variables are correlated among themselves (Table 5-2 shows a part of correlation matrix). The eight factors are uncorrelated (Table 5-10). For analytical work, uncorrelated variables are of great value since most of the statistical tests of significance are based on independent (i.e., uncorrelated in case of normal) variables. Moreover, the correlations between any pair of observables (for the exploratory subsample part of data) can be reproduced by a linear combination of factor loadings (see **Ref 5-1**). This completes our objective of data reduction without sacrificing information.

Table 5-10. Pearson Correlation Coefficients Between Factors for the Exploratory Subsample[a]

Factor	F1	F2	F3	F4	F5	F6	F7	F8
F1	1.0000 (92) P=*****	.0033 (92) P= .487	-.0451 (92) P= .336	-.0229 (92) P= .414	-.0113 (92) P= .457	-.0182 (92) P= .432	.0243 (92) P= .409	.0172 (92) P= .435
F2	.0033 (92) P= .487	1.0000 (92) P=*****	.0047 (92) P= .482	.0100 (92) P= .462	.0129 (92) P= .452	-.0124 (92) P= .453	-.0060 (92) P= .478	-.0092 (92) P= .465
F3	-.0451 (92) P= .335	.0047 (92) P= .482	1.0000 (92) P=*****	-.0384 (92) P= .358	-.0612 (92) P= .281	-.0760 (92) P= .236	.0076 (92) P= .471	.0085 (92) P= .468
F4	-.0229 (92) P= .414	.0100 (92) P= .462	-.0384 (92) P= .356	1.0000 (92) P=*****	.0347 (92) P= .371	.0395 (92) P= .354	.0245 (92) P= .408	-.0233 (92) P= .413
F5	-.0113 (92) P= .457	.0129 (92) P= .452	-.0612 (92) P= .281	.0347 (92) P= .371	1.0000 (92) P=*****	-.0046 (92) P= .482	.0002 (92) P= .499	.0619 (92) P= .279
F6	-.0182 (92) P= .432	-.0124 (92) P= .453	-.0750 (92) P= .236	.0395 (92) P= .354	-.0046 (92) P= .482	1.0000 (92) P=*****	-.0077 (92) P= .471	-.0086 (92) P= .468
F7	.0243 (92) P= .409	-.0060 (92) P= .478	.0076 (92) P= .471	.0245 (92) P= .408	.0002 (92) P= .499	-.0077 (92) P= .471	1.0000 (92) P=*****	-.0041 (92) P= .485
F8	.0172 (92) P= .435	-.0092 (92) P= .465	.0085 (92) P= .468	-.0233 (92) P= .413	.0619 (92) P= .279	-.0086 (92) P= .468	-.0041 (92) P= .485	1.0000 (92) P=*****

[a]Values given are the correlation coefficient, the number of data points used (in parentheses), and the significance level of the correlation coefficients (the so-called P-value).

5-8. NEXT STEPS FOR ANALYSIS OF REDUNDANCY. Some of the desirable next steps for analyzing redundancy are presented in Table 5-11. Naturally the preliminary calculations presented earlier in this chapter should be redone as the CDES contract provides revised data base values. Also, what data subsets should be used depends in part on how the World War II anomaly is resolved. Although a number of alternative techniques are available, several of which are listed in Table 5-11, this chapter used factor analysis for analyzing redundancy because:

a. Factor analysis is a tested and relatively objective method that yields reproducible results, and that has been widely used in the social sciences.

b. It is useful in cases where the variables are highly correlated, as is the case with the HERO data.

c. Computer programs for factor analysis are available at CAA in standard statistical computer program packages (Ref 5-1).

d. Factor analysis is well-suited to an exploratory activity because it requires little preparation and analysis effort.

However, a word of caution is in order here. Although one of the assumptions in factor analysis is that the variables are continuous, nearly all of the variables in this chapter are discrete rather than continuous. Therefore, the results obtained here are indicative rather than rigorously established. Consequently, future efforts should consider applying more flexible and powerful statistical techniques for redundancy analysis.

The alternatives to factor analysis have their own strengths and weaknesses. For example, cluster analysis (Ref 5-7), projection pursuit (Ref 5-8), and (non-linear) transformations for maximum correlation (Ref 5-9) are relatively recently-developed methods. Computer programs for implementing them are not yet offered in standard statistical computer program packages. Because CAA has little prior experience with these programs, a significant start-up cost may be required to bring them on-line at CAA and to learn how to use them effectively. Also, some of the alternatives involve fairly advanced statistical methods and require much more preparation and analysis effort than factor analysis. Examples--in addition to the relatively new methods mentioned above--are discriminant analysis (Ref 5-10), canonical correlation (Ref 5-11), and stepwise regression (Ref 5-12). Informal examination of the correlation matrix (Ref 5-13), while it requires very little additional preparation, provides results that are far more subjective and dependent on the skill and personal idiosyncracies of the analyst than are factor analysis results. Naturally, whatever technique or combination of techniques may be used for redundancy analysis, the results obtained should be documented in an appropriate form.

Table 5-11. Next Steps for Analysis of Redundancy

1. Revise the Preliminary Calculations as CDES Results Become Available.

2. What data subsets to use hinges on resolution of the World War II Anomaly.

3. Explore Alternative Techniques for Analyzing Redundancy.

   a. Informal Examination of the Correlation matrix

   b. Cluster Analysis

   c. Projection Pursuit

   d. Discriminant Analysis

   e. Canonical Correlation Analysis

   f. Stepwise Regression Analysis

   g. Transformations for Maximum Correlation

4. Document the Results Obtained

5-9. CONCLUDING OBSERVATIONS ON THE ANALYSIS OF REDUNDANCY

a. The problem of determining what underlying or "basic" factors undergird a given set of observations has vexed statisticians, scientists and philosophers for thousands of years. Factor analysis is currently reputed to be one of the more frequently used techniques, and we have applied it to 29 variables from the HERO data base. The results indicate that the original 29 variables can be replaced in future analyses by 8 new factors that are uncorrelated (i.e., not redundant) with practically no loss of information (in the technical sense).

b. Nevertheless, the 8 new factors produced by the principal factors method may not be as intuitively clear as the original 29 were. Also, the present analysis intermingles variables from the original HERO data base Tables 4 and 6, although there may be good reasons to keep them separate. Moreover, we have applied factor analysis methods to discrete data even though the method technically requires the data to be continuous. Accordingly, the present exploratory effort at redundancy analysis must be used with caution, and future analyses should consider alternative approaches.

CHAPTER 6

TEST OF A BREAKPOINT HYPOTHESIS

6-1. INTRODUCTION

a. The purpose of this chapter is to address the validity of a breakpoint-type hypothesis for determining the terminal status of a land battle. The version of the breakpoint hypothesis used is a moderate simplification of the ones frequently used to determine when and how to terminate simulated combat for various types of combat models, such as those used in wargames, computer simulations, and the like. It is as follows:

- Each side selects independently a breakpoint from a distribution of such breakpoints and gives up the battle when its casualty fraction reaches its breakpoint.
- These breakpoint distribution curves are generally applicable.
- The casualty fractions of the forces are deterministically and monotonically related to each other.

Some of the major theoretical implications of this breakpoint hypothesis are quantitatively compared against casualty-fraction distribution data from the HERO data base.

b. The principal finding is that the above breakpoint hypothesis is contradicted by the available historical data. However, the task of devising a theory that satisfactorily accounts for the available data is not within the scope of this paper. Until a better theoretical explanation of the battle termination process becomes available, the soundness of models of combat such as war games and computer simulations that make essential use of breakpoint hypotheses is suspect.

c. The breakpoint hypothesis has been tested previously using the CORG and the BWS data bases **(Ref 6-1)**. The results obtained here, using the HERO data base, support and confirm these earlier results. Much of the material in this chapter is based on Ref 6-1 and extracts from it are used liberally in this paragraph and also in paragraphs 6-2 and 6-3.

6-2. ORIENTATION

a. Consider two opposing forces engaged in a land battle. As the engagement continues, both sides will suffer casualties. Eventually, the battle will end. At the termination of the engagement, the situation may be one of the following:

- One side has, for all practical purposes, been annihilated, leaving its opponent in control of the battlefield.

- One side surrenders and submits to the will of its opponent, who thereby acquires control of the battlefield.

- Neither side has surrendered or been annihilated, but one of them has disengaged and either has withdrawn or is in the process of withdrawing from the area, leaving its opponent rather clearly in control of the battlefield.

- Neither side has surrendered or been annihilated, but both sides have disengaged their forces, and both sides either have withdrawn or are in the process of withdrawing their forces from the area. The withdrawal is mutual, and it is impossible, or at any rate a very difficult and controversial matter, to assert that either side has practically exclusive control of the battlefield.

b. This list of possibilities excludes a situation that occasionally occurs, in which both sides have disengaged their forces, but neither side appears ready to leave the field. Sporadic skirmishes may be taking place along the line of demarcation. (Typically, this sort of situation occurs when a defensive force is reluctant to leave a strong defensive position in the presence of a relatively stronger enemy who considers that an immediate assault would not be worth the probable losses.) These conditions evidently describe a kind of unstable standoff that will eventually resolve itself either into a renewal of the engagement or into one of the four kinds of termination described earlier, so we will view the standoff case as a temporary pause or lull in hostilities, rather than as a conclusion of the engagement.

c. Of the four terminal situations listed, the second and third, where there is a fairly clear-cut victor, seem to be the most common. Possession of the battlefield seems to be a generally accepted criterion of victory in the battle. There are cases in which the battle loser has imposed a serious strategic cost on the tactical battlefield winner. The "Pyrrhic" victory (Battle of Asculum, 279 B.C.) is a famous example of a tactical victory obtained at a heavy strategic loss. Annihilation is quite rare except in circumstances where retreat is impossible (as may occur, for example, in sieges or in island campaigns). Even where retreat is out of the question, a defender whose position is deteriorating will normally surrender rather than fight to the last man. Mutual withdrawal, with its inconclusive outcome, although more frequent than annihilation, is still a relatively rare occurrence. In general, a weakening side will prefer to withdraw and abandon the field rather than surrender to its opponent, and (if withdrawal is not feasible) will usually prefer to surrender at some casualty level short of 100 percent total annihilation.

d. A so-called "break curve" is a device often used to model the inclination of a weakening force to discontinue the engagement by acknowledging defeat and either withdrawing (if it can) or surrendering. It is a curve that purports to show the probability that a force will discontinue the engagement as a function of the casualty fraction that it has sustained. Figure 6-1 shows a hypothetical break curve. A break curve is often used in combat models as follows. At or before the beginning of a simulated engagement, a sample casualty-fraction value for each side is drawn from the distribution of such values defined by an appropriate break curve. The values so selected are called the "breakpoints" for the two sides. Then, as the engagement progresses, both sides are considered to be engaged in a contest for dominance until one of them accumulates enough casualties to equal or exceed its preselected breakpoint. At that point, the side whose preselected breakpoint has been reached is said to "break," meaning that it is presumed to discontinue or "break off" its attempts to dominate the opposing side. Thus, the side that breaks is considered by the rules of this particular model to lose the battle.

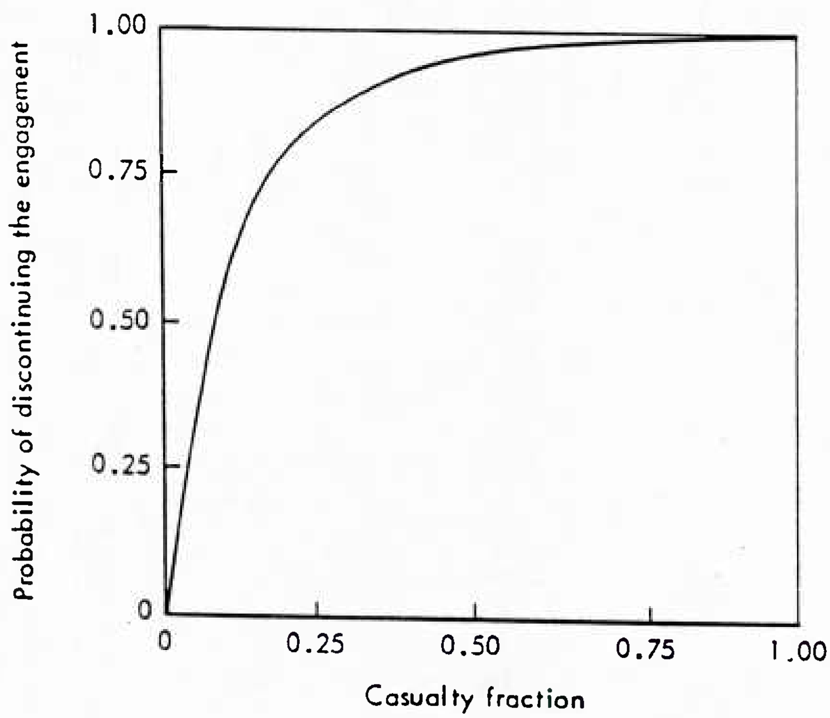


Figure 6-1. Hypothetical Break Curve

e. Break curves of the sort just described are presented in **Ref 6-2** (paragraph 15, Appendix IV) and elsewhere. Frequently, application of the break curves is simplified by assuming that breaks occur deterministically. The break curve approach described above can be adjusted to this case by taking the break curve to be a step function with a vertical rise of unity at the deterministic breakpoint, as indicated in Figure 6-2. This special type of break curve will be called a deterministic break curve. Perhaps the most common type of break curve proposed is of the deterministic type.

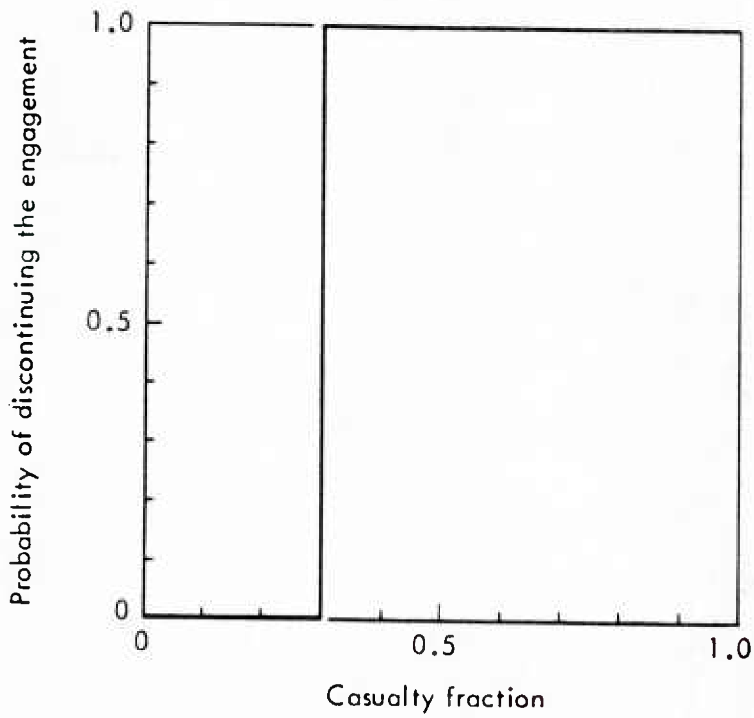


Figure 6-2. A Deterministic Break Curve

f. Objections to the validity of deterministic break curves as descriptors of combat behavior have been voiced from time to time. For example, according to **Ref 6-3**, "The statement that a unit can be considered no longer combat effective when it has suffered a specific casualty percentage is a gross oversimplification not supported by combat data." The collection of casualty data included in Appendix F of **Ref 6-1** and in paragraph 3-7 of this paper confirms this conclusion. **Ref 6-3** showed that a deterministic-type break curve is not generally applicable to the observed behavior of combat units but did not analyze the validity of the more general type of break curve illustrated in Figure 6-1. At present, the validity of the more general type of break curve seems to be a controversial matter. On one hand, some analysts have proposed their use for wargaming, maneuver control, and similar purposes, as noted earlier. Other analysts have designed simulations using the simpler and more specialized deterministic break curves, despite **Ref 6-3's** objections to their merit, and so by implication have embraced the basic philosophy that unit behavior is representable by some type of break curve.

g. On the other hand, some analysts have grave misgivings about the validity of break curves--even while they may, on occasion, use them for lack of anything better. Some of the objections raised against the use of break curves are discussed below. Most of them can be characterized as suggesting that some other factor or factors than simply the current casualty level of a force influence the break behavior of the force. Frequently these other factors are proposed as considerations supplementary to, rather than as replacements for, the casualty-level criteria. This suggests that the casualty level is often thought of as a sort of "core" consideration that may be modified in particular situations by some of these additional considerations.

h. For example, it is sometimes suggested that the casualty rate, as well as the casualty level, influences the behavior of a force. Other considerations include the level of training and battle experience of the troops, the influence of inclement weather or other unusual environmental stress, the importance of the mission, troop morale, the quality of leadership, the degree of knowledge and intelligence of the enemy's situation and intentions, the perceived vigor of the enemy opposition, the scale of friendly fire support and troop reinforcement, the logistical supply situation, and the availability of good communications with other friendly units. Many of the considerations that impinge on the intuitive plausibility of the break curve approach are carefully discussed in **Ref 6-3**. We do not intend to pursue the extent to which the break curve model's "face validity" is affected by these plausibility arguments, since we will confront our model with empirical data in order to determine its validity.

i. However, there is one further objection that has been raised against the break curve approach that needs to be discussed in somewhat more detail. This is the observation that each side in an actual battle surely considers the progress of the battle and continually assesses its own situation relative to that of its opponent, rather than being governed solely by its own condition. In this view, each side conducts itself according to the results of a dynamic decision process lasting throughout the battle rather than preselecting a specific breakpoint, as is done in the conventional application of break curves to war games, simulations, and field maneuvers. That the objection is not always relevant is demonstrated by the discussion in Appendix C of **Ref 6-1**, where it is shown how most types of continuous decision processes can be subsumed under the break curve paradigm without losing any generality. The key assumption in such derivations is the supposition that each side, while it may decide continually whether to continue the engagement or not, bases the decision solely on its own current casualty fraction. Similar derivations of break curves from dynamic decision processes have been given in **Refs 6-4, 6-5, and 6-6**. In none of these derivations is the possibility explicitly considered that one side's breakpoint may depend on the casualty level of its foe. Thus, it seems that in order for the objection raised earlier (that break curves fail to reflect the dynamic decision processes actually taking place in combat) to retain its validity it must also be supposed as a minimum that one side's breakpoint distribution depends on the other side's casualty level.

j. In addition to the conceptual issues discussed above, there are several practical problems in assessing the validity of breakpoint assumptions. These stem from the kind of empirical evidence that is more-or-less readily available for comparisons with the model. First, the recoverable data are essentially limited to estimates of the attacker and defender initial troop strength, of the total losses* on each side, and (occasionally) of the temporal duration of the battle, together with a narrative account of the action and a historical judgment either awarding the victory to one side or the other or declaring the outcome "indecisive." Second, the criteria for assessing casualties may vary among battle descriptions from very broad to highly restrictive. Third, there is often much scope for human error and/or capriciousness in selecting the forces to be included in establishing troop strength or casualties, as well as in arriving at an accurate inventory of these quantities. These problems are noted and discussed a bit further in **Ref 6-7**, but no solution to them (short of a detailed and thorough reexamination of the original historical records) is in evidence. These problems make enlarging the sample size a tedious, time-consuming, and very expensive task. Such is the nature of the basic data at our disposal.

k. To the above difficulties yet another must be added--namely that the attrition dynamics intervene between the break curve and the observed battle outcome and force ratio. That is, after breakpoints are established, parallel casualty assessments for each side must be made in order to determine the final outcome and casualty fractions. Consequently, it is clearly incorrect to establish a break curve by simply plotting the cumulative fraction of battles that terminated before various casualty-fraction levels were sustained. A correct analysis of the relation of observed casualty-fraction distributions and break curves is given in Chapter II of **Ref 6-1.**

l. The next paragraph summarizes the results of that analysis, states the breakpoint hypothesis that will be tested, and describes the method used to test it.

---

*Not necessarily only those inflicted prior to reaching a breakpoint. In some cases, the historically reported casualties may have occurred after the break. For example, routs sometimes degenerate into massacres, and on occasion troops that have surrendered may have been slain.

6-3. STATEMENT OF THE BREAKPOINT HYPOTHESIS. The breakpoint model considered here is founded on the following postulates. The ensuing development requires each of the assumptions made, as well as some additional ones that will be introduced as we go along.

a. **Postulate A.** Termination of a battle can be considered as governed by the following mechanism, or one that gives the same results: Prior to the battle, each side independently, and at random, selects a casualty-fraction value (breakpoint) from some distribution of casualty fractions. When either side experiences a casualty fraction equal to its preselected breakpoint, the battle terminates with a loss to the side that "broke."*

b. **Postulate B.** The breakpoint distributions (break curves) mentioned above are generally applicable. That is, they are the same for all battles, irrespective of the size of forces involved or when, where, by whom, or with what the battle was fought.

c. **Comments on Postulates A and B.**

(1) Postulates A and B are introduced because they are in fact the way break curves are used in many war games and combat simulations. Postulate B can be tested by various groupings of empirical battle data, and also makes explicit an assumption that is often overlooked. Postulate B is, to a large extent, provisional, in that we may modify it if the

---

*In employing the casualty-fraction value as the key parameter value, there is a tacit assumption that the battle is fought to its conclusion with the forces on hand at the start, since this provides a well-defined base for establishing the casualty fraction. If reinforcements occur during the battle, then it is necessary to have some further rules about how to determine the casualty fraction. For example, Clark (**Ref 6-1**) computes distinct casualty-fraction values two ways: (1) cumulative casualties from start of engagement per troop at the start, and (2) the difference, cumulative casualties less cumulative replacements, per troop at the start. In other contexts, reinforcements are often modeled in one of two extremes, i.e., either they are assumed to have a negligible impact on the situation and ignored (perhaps with some rationalization to the effect that they arrived too late to affect the outcome), or they are lumped with the initial forces and so are counted as being fully effective throughout the battle. In this paper, we shall take the initial forces given in the references consulted as the base for determining casualty fractions.

empirical data warrant it. It is certainly a rather strong and perhaps controversial postulate, once it is clearly stated. However, it is hoped that it may be testable, whereas the opposite tack of assuming that every battle fought has its own special break curves which depend on the unique circumstances surrounding the particular battle is not likely to lead to a theory that can be compared with such data as are available.

(2) While data from which accurate curves may be drawn are hard to come by, there is no other reason for restricting the method to a single break curve. In principle, the appropriate break curve could be made to depend on any condition that could be known at the time the break curve is sampled, such as whether the force is initially attacking or defending, its state of training, experience, morale, physical weariness, etc. We will not pursue this possibility here. The approach adopted is in keeping with the spirit of Richardson's Principle to the effect that "formulae are not to be complicated without good evidence" **(Ref 6-8, p. xliv).**

d. Notational Conventions

(1) Some notation needs to be introduced at this point (also see the Glossary). Let FX(t) and FY(t) be the fraction of casualties for side X (attacker) and side Y (defender) as of time t after the start of the battle. Let LX and LY be the breakpoints or casualty-fraction threshold values for the attacker (side X) and defender (side Y), respectively. Let FX and FY be the fraction of casualties sustained by the attacker and the defender during the whole course of the engagement.

(2) By virtue of the breakpoint hypothesis, LX and LY are random variables with appropriate distributions. At the conclusion of the battle, either FX or FY is equal to its corresponding breakpoint, while the other is less. Thus, we have either FX is less than LX and FY = LY (in which case the attacker wins), or FX = LX and FY is less than LY (in which case the defender wins). In either case, both FX(t) is less than LX and FY(t) is less than LY hold for all times t from the onset of the battle to its conclusion, i.e., for t at least 0 and t not exceeding T. With this notation, we can introduce postulate C.

e. Postulate C. The losses, and hence equivalently the casualty fractions, of the forces are deterministically and monotonically related to each other. That is, there is a monotonically increasing function, $\Psi(\cdot)$, such that

$$FX(t) = \Psi\ (FY(t)) \quad ,$$

for all t greater than zero and less than T.

f. Concluding Observations on Postulates A, B and C. It would be of interest to consider the effect of assuming nondeterministic and/or nonmonotonic relationships between the two casualty fractions although such an investigation is not within the scope of this analysis. The assumption made here is a generalization of that made by Weiss*, who assumes that the casualty fractions are proportional to each other **(see Ref 6-6, p. 776)**, i.e., that there is a "fractional exchange ratio," R, such that

$$FX(t) = RFY(t).$$

This is equivalent (provided, of course, that R is greater than zero) to the special case of $\Psi(u) = Ru$. At a later point in the argument, we will find it useful to introduce particular forms of the function Ψ. The real reason for assuming Ψ to be strictly monotonic is to assure that it will have a uniquely definable inverse, Ψ^{-1}, whose role is made clear by ensuing developments.

6-4. METHOD FOR TESTING THE BREAKPOINT HYPOTHESIS

a. Selected Consequences of the Breakpoint Hypothesis. In Ref 6-1 it is shown that the breakpoint hypothesis stated in paragraph 6-3 has the following logical consequences:

$$P(FX \le s \,/\, ATKWIN) = P(FY \le \Psi^{-1}(s) \,/\, ATKWIN) \qquad (6\text{-}1.1)$$

$$P(FY \le s \,/\, DEFWIN) = P(FX \le \Psi(s) \,/\, DEFWIN), \qquad (6\text{-}1.2)$$

where ATKWIN means the attacker wins (i.e., WINA = +1) and DEFWIN means the defender wins (i.e., WINA = -1).

---

*The details of Weiss' subsequent development diverge from ours in that he introduces a model of break behavior in terms of a continual, but mutually independent evaluation of current status by each side. However, as was noted earlier, the approach presented here applies to this case also, once the break curves for each side have been derived from the dynamic model of each side's decision behavior **(Ref 6-1, Appendix C)**.

b. Now suppose that we form a graphical plot of the observed or empirical casualty fractions for a collection of battles that were won by the attacker. A hypothetical plot is shown in Figure 6-3, on which the dashed lines indicate how, using equation (6-1.1), the value of $\Psi^{-1}(s)$ can be graphically read off this plot. Reference 6-1 gives the mathematical justification for this procedure. An exactly analogous procedure applied to the corresponding plot for battles won by the defender will yield the value of $\Psi(s)$. By repeating the process for several values of s and interpolating, it is thus possible to determine suitable approximations to both of the functions and Ψ and Ψ^{-1}.

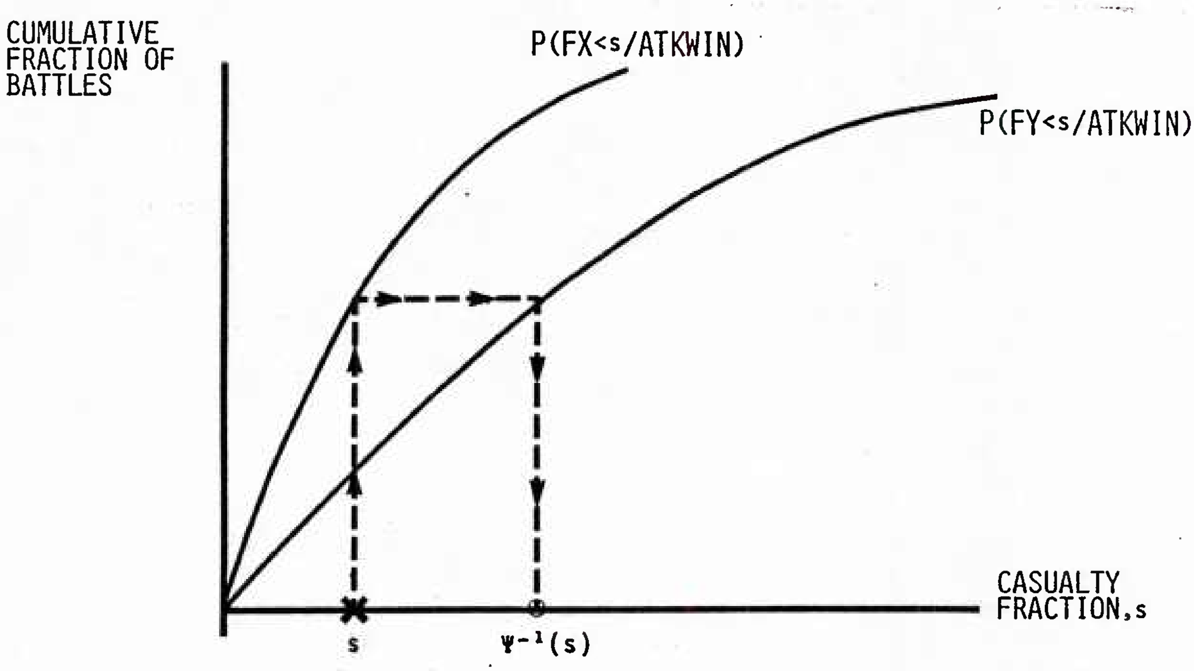


Figure 6-3. Construction of the Ψ^{-1} Function

c. Having determined Ψ and Ψ^{-1} by the graphical procedure just described, we may plot these functions on a graph and see whether or not they obey the necessary mathematical relationship between inverse functions, that is, whether or not Ψ is a reflection of Ψ^{-1} in the 45-degree line through the origin, as illustrated in Figure 6-4. If Ψ and Ψ^{-1} obey the inverse functional relationship, then this would tend to support the breakpoint hypothesis. If Ψ and Ψ^{-1} do not obey the necessary mathematical relationship between inverse functions, then the breakpoint hypothesis would be definitely disproven. We shall carry out just such a test in the next paragraph.

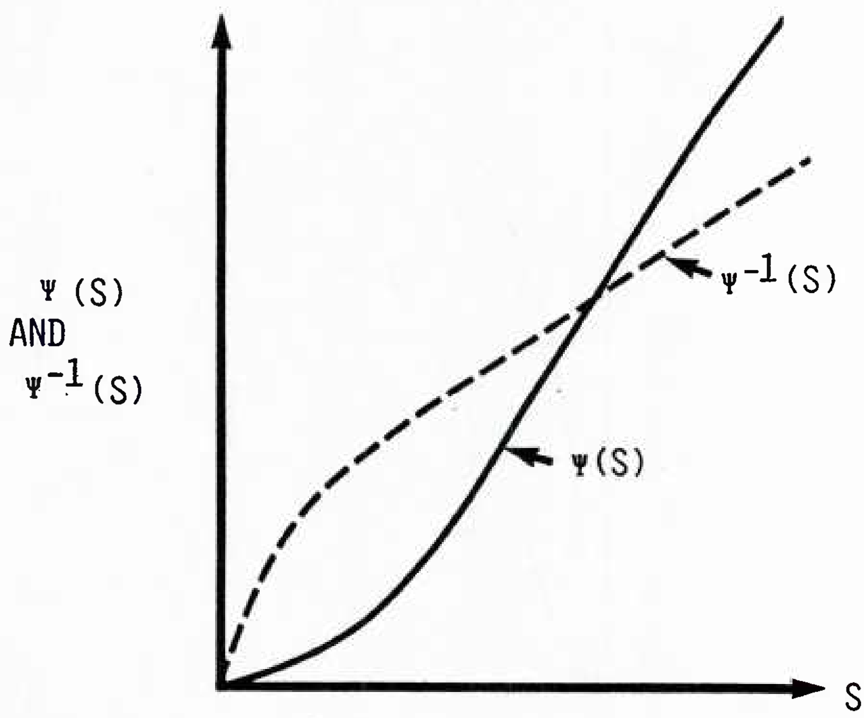


Figure 6-4. Test the Logical Consequence of the Breakpoint Hypothesis: Are Ψ and Ψ^{-1} Inverse Functions?

6-5. TEST OF THE BREAKPOINT HYPOTHESIS. The empirical data needed to test the hypothesis are the distributions of casualty fractions conditioned on who wins, as symbolically expressed by equations (6-1). Curves of this type, determined by the HERO data base values, are shown in Figures 6-5 and 6-6. They were generated using the non-WWII data subset with draws counted as defender wins. The empirical Ψ and Ψ^{-1} functions generated from these casualty fraction distributions are shown in Figure 6-7. Clearly, the empirical Ψ and Ψ^{-1} functions are <u>not</u> related as inverse functional relationships (see Figure 6-4). Consequently, the breakpoint hypothesis stated in paragraph 6-3 cannot be correct--at least one of its three postulates must be wrong.

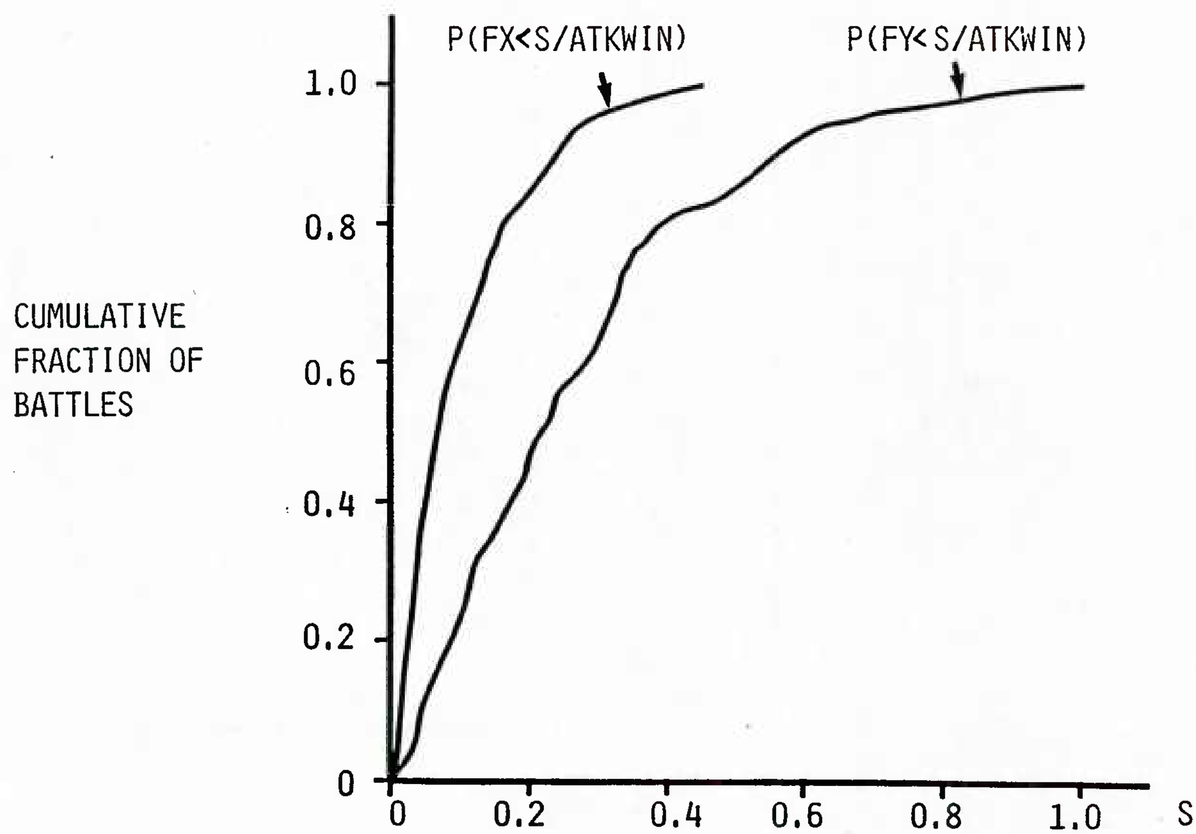


Figure 6-5. Distribution of Casualty Fractions When the Attacker Wins

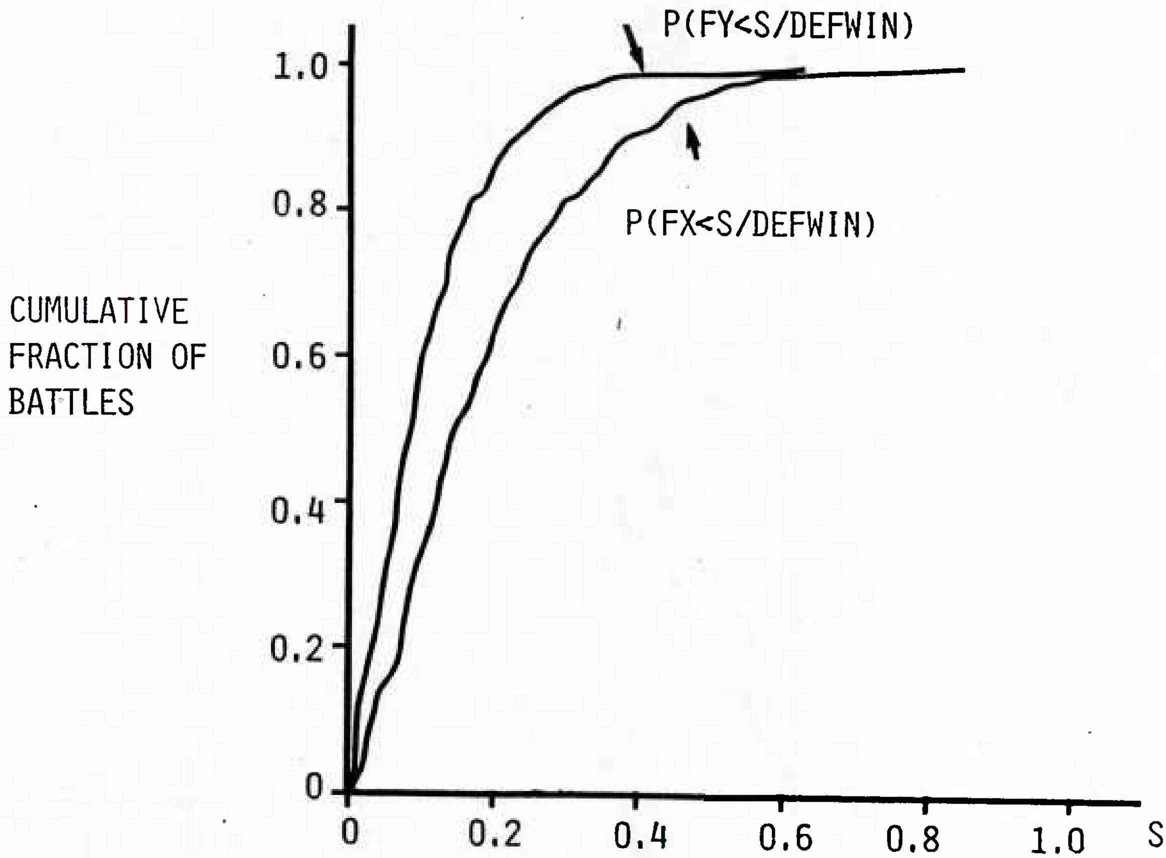


Figure 6-6. Distribution of Casualty Fractions When the Defender Wins

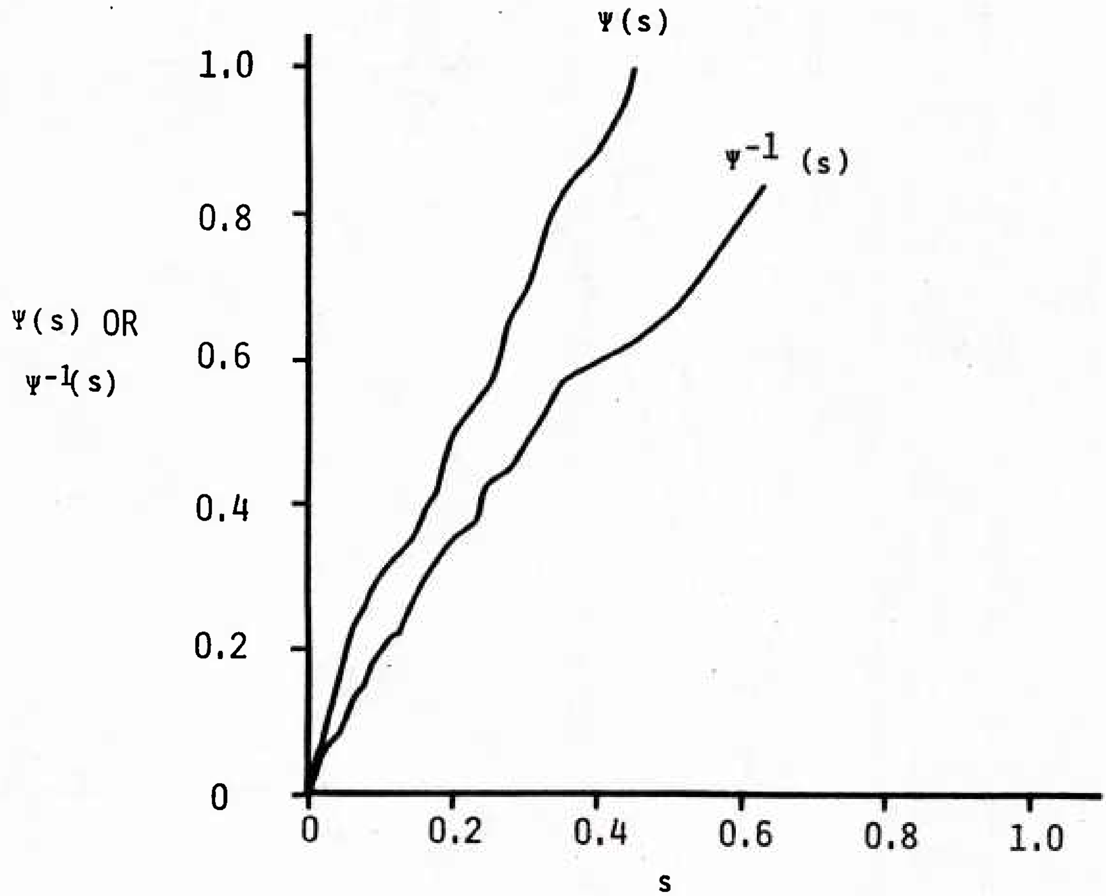


Figure 6-7. Comparison of Empirical Ψ (s) and Ψ^{-1}(s) Functions

6-6. NEXT STEPS FOR HYPOTHESIS TESTING. Some of the next steps for hypothesis testing are given in Table 6-1. The Ψ and Ψ^{-1} functions should be recomputed when the CDES contract provides revised data on particular battles. What data subsets should be used in computing these functions depends in part upon how the WWII anomaly is resolved. Several hypotheses in addition to those concerned with breakpoints can be considered and it may be possible to design appropriate ways of testing some of them by using data from the HERO data base. Finally, any results that are obtained should be documented in an appropriate form.

Table 6-1. Next Steps for Hypothesis Testing

1. Revise the preliminary calculations as CDES results become available.
2. What data subsets to use hinges on resolution of the WWII anomaly.
3. What other hypotheses can be tested?
   a. Three-to-one force ratio for successful attack?
   b. Advance rate increases with force ratio?
   c. Air support significantly enhances P(WIN)?
   d. Fortifications significantly raise defender's P(WIN)?
   e. Casualty exchange ratio improves with force ratio?
   f. Fractional exchange ratio has not changed much over the years?
   g. Are EPS and ADV constant with respect to time during a battle? Can they be estimated accurately from an earlier portion of a battle? Can they be predicted before the battle is joined?
   h. Any evidence for attrition laws other than the square law?
   i. What is the evidence for "Osipov's Law," that losses are inversely proportional to the square roots of the initial strengths? Is the attrition fraction (or rate) lower for large forces than for smaller ones?
   j. Are EPS and ADV independent?
   k. Does EPS and/or the casualty rate increase with advances in weapons technology?
   l. Is EPS directly proportional to the duration of battle (so that LAMBDA is nearly constant), or is LAMBDA inversely proportional to the duration of battle (so that EPS is nearly constant)?
   m. Do battles with ADV near zero tend to last longer and/or to be more bitter than those with high or low ADV values? Is this true when both sides can fairly readily break off the engagement?
   n. In the US Civil War, did declining morale and equipment cause losses in battles, or did the battle losses cause the decline in morale and equipment? Is there a secular trend of ADV with respect to battle date during the Civil War?
   o. Is there a critical force ratio at which a side engaged in a battle will "break?"
4. Document the findings.

6-7. CONCLUDING OBSERVATIONS ON HYPOTHESIS TESTING

a. We have presented a test of a breakpoint hypothesis to illustrate the potential of hypothesis testing as a method for using combat data to study wargaming issues. This work may also serve as instructive example for future efforts at hypothesis testing.

b. The particular breakpoint hypothesis considered was shown to be false. This result casts doubt on the validity of the break curves constantly used in wargames.

c. Devising a satisfactory theory of victory in tactical operations was not within the scope of the effort reported in this paper.

CHAPTER 7

OTHER ANALYSES

7-1. INTRODUCTION. A number of important and interesting analyses, which could not be addressed within the scope of the effort described in this paper, are planned for future work. They are called for by the CHASE study directive's objectives and essential elements of analysis (EEA), or are part of the CHASE study plan. This chapter describes the objectives, scope, and next steps for future analyses dealing with (1) rates of advance, (2) the influence of air support, and (3) long-term trends. However, the future investigation of other important issues that may arise is not precluded, even though they may not fit neatly within the categories mentioned earlier. Rather, issues will be addressed in priority order after considering the following factors bearing on their priority (listed roughly in order of importance):

a. Relative current interest or importance of the issue for military operations research, concept formulation, wargaming, and studies and analyses.

b. Relative prospects of obtaining useful and relevant results from an analysis of the types of information provided by the HERO, CORG, and BWS data bases.

c. Relative benefits (in the form of computer programs, necessary preliminary steps, insights, etc.) to subsequent phases of CHASE from conducting the investigation at this time.

d. Relative ease of performing the analysis.

7-2. RATES OF ADVANCE

a. Orientation. One of the CHASE study's EEAs is "What can be said about the factors influencing rates of advance in land combat?" In addressing this EEA it must be recognized that the HERO data base deals with battles rather than with theater operations, campaigns, or wars. Consequently, its information on distances and rates of advance are those for forces that are fully engaged in combat. For example, the HERO data base does not provide data on the average rates of advance in campaigns or in unopposed operations. In addition, careful attention must be given to the following definitions used in the original HERO data base:

(1) Attacker (NAMA). "That military force which, at the beginning or in the first phase of an engagement, initiates and sustains significant offensive action against its opponent."

(2) **Duration (T).** "The extent of time, expressed in number of days, during which an engagement takes place." In the HERO data base, a portion of a day is considered a full day, except in cases of overnight engagements in which significant combat began in late afternoon or evening and was concluded before noon on the following day. In such cases the engagements are considered one-day engagements, since the duration was less than twenty-four hours.

(3) **Distance Advanced (KPDA).** "That distance, in kilometers, from the line of departure to the farthest point reached by significant maneuver elements of the attacking force, measured along the axis of advance". (NOTE: The values actually published in the data base are rates of advance in km/day, rather than distances in km).

These definitions sharply limit the derivable conclusions on rates of advance. One important problem is that the KPDA values reflect only the position of forces at the end of the battle. For example, negligible or zero values of KPDA may represent a practically stationary line of contact, an unsuccessful slight penetration followed by a slight retreat by the attacker, an unsuccessful deep penetration followed by a counterattack that restored the line and then drove deeply into the attacker's initial position but was eventually repulsed, leaving both sides close to their initial positions, etc. Also, a modest value of KPDA may represent a successful permanent advance by the attacker, an initial attacker success that was halted by a sharp counterattack, a planned defender withdrawal followed by an envelopment that totally defeats the attacker, etc. Moreover, since the maximum attacker penetration will not always correspond to the end of the battle, dividing the attacker's penetration distance by the total battle duration introduces a systematic bias toward lower advance rates than were actually attained. A further systematic bias tending to make the reported rates of advance smaller than they are in reality is introduced by the use of whole days for durations, rather than more exact values. In sum, the HERO data are not well suited to an analysis of rates of advance that would be comparable to previous work **(i.e., Refs 7-1 through 7-12).**

b. Next Steps. As noted in Chapter 2, refined time duration data are being provided under the CDES Contract, and this will eliminate the bias caused by the use of whole days rather than true duration data. However, it will not affect the bias introduced by assuming that the maximum penetration occurred at the end of the battle. Also, what battles to include depends in part on how the WWII anomaly is resolved. In view of the observations made in paragraph 7-2a, the foreseeable next steps are limited to inquiring whether the rates or distances advanced by the attacker depend in an important way on such variables as force ratio, level of air support, casualty exchange ratio, fraction exchange ratio, ADV, EPS, weather, terrain, etc. The findings will then have to be interpreted and documented in suitable form.

7-3. AIR SUPPORT

a. Orientation. One of the EEAs is "Can the historical influence of air support on the outcome of land battles be quantified"? Since neither the CORG nor the BWS data bases contain information on air support, the analysis will necessarily be limited to the data in the HERO data base. The HERO data base gives, for some battles of the WWI era on, judgments of which side had air superiority, how much air superiority favored one side or the other, the number of close air support sorties flown by each side, and the number of these aircraft that were lost in combat on each side. However, this information is not given for all battles. Also, except for the number of aircraft lost, the HERO data base provides no information on the local air defense capabilities of the two sides. These conditions limit the kinds of analyses than can usefully be accomplished.

b. Next Steps. Since most of the battles with air support data are from the post-1940 era, the manner in which the WWII anomaly is resolved will significantly affect the air support analysis. However, it may be possible to determine whether the general level of adjudged air superiority and/or the number of close air support sorties flown by each side significantly:

(1) Increases the probability of winning,

(2) Accelerates the rate of advance, or increases the depth of penetration,

(3) Shortens battle duration, increases battle intensity (LAMDA), or alters bitterness (EPS),

(4) Improves the casualty exchange ratio (CER) or other variables such as FER, ADV, ACHA, or ACHD,

(5) Heightens tactical surprise, and

(6) Influences other factors to be determined.

Of course, any results must be interpreted and documented in suitable form.

7-4. LONG-TERM TRENDS

a. Orientation. One of the CHASE Study's EEAs is "What long-term trends can be detected in historical combat data?" What trends are perceived depends in part on the timespan covered by the data. The CORG data base has fewer battles and spans about the same time period as does the HERO data base, but includes a few battles from ancient times. The BWS data base has more battles but spans only the time period 1618-1905, as compared to the HERO data base's 1600-1973 time span. Both the CORG and BWS data bases will probably supplement usefully the HERO data base for some aspects of long-term trend analysis.

b. Prior Work. Several publications **(Refs 7-13 thru 7-18)** describe some earlier work on trends. Of these, only Reference 7-15 is applicable to battles and engagements. The others deal with wars and campaigns, which transcend the information given in the HERO, BWS, and CORG data bases. However, Reference 7-15 does not give sample sizes, does not provide the raw data, and its published findings cannot be traced to the original data sources. Nevertheless, its conclusions suggest the type of information on trends that may be obtainable:

"A quantitative analysis of the major wars of the last 250 years shows increasing trends in the following:

"**(a)** The magnitude of wars as measured either by the total numbers mobilized or the average effective strengths of the armies. More countries tend to take part in modern wars and greater proportions of their populations become involved.

"**(b)** The average size of armies in battle.

"**(c)** The cost of wars, measured either by the actual cost or the proportion of their national incomes to participating nations spent on them.

"**(d)** The deadliness of wars, measured by the total number of casualties incurred in them.

"**(e)** The proportion of artillery and supply troops in armies.

"**(f)** The proportion of casualties caused by fragmentation weapons.

"**(g)** The average small arms and artillery firepower of armies per 1,000 combat troops.

"**(h)** The average cost in ammunition of causing casualties.

"**(i)** The average range at which casualties are caused.

"**(j)** The daily supply requirements per man.

"A detailed analysis of these trends, together with a consideration of present and probable future developments in warfare is necessary before an estimate can be made of their probable continuance.

"The ratios of the total numbers of casualties to the total numbers mobilized in the various wars has tended to increase with the length of the wars, but the average probability of any soldier becoming a casualty at any time during any one of those wars has not shown a large variance or any marked trend.

"As the daily expenditure of ammunition by troops has increased, the cost in ammunition of causing casualties has tended to show a similar increase. The introduction of more effective means of attack during this period have been counteracted by improved defensive measures which have proved more or less as effective."

c. **Next Steps**

(1) Detection of Trends. After revising the data on the basis of the CDES contract results and deciding how to handle the WWII anomaly, the data can be examined for evidence of the following hypothesized long-term trends:

(a) Increasing casualty fractions.

(b) Larger forces.

(c) Wider fronts.

(d) Decreasing linear troop density.

(e) Longer battle durations.

(f) Lower casualty rates.

(g) Long-period oscillation between offensive and defensive preponderance.

(h) Changes with respect to battle date in such variables as FR, ADV, EPS, T, LAMBDA, FER, CER, XO, YO, P(WIN), etc.

(2) Interpretation of Trends. Any trends discovered above must, of course, be interpreted and documented in suitable form. The significance of trends may be clarified by relating them to parallel trends in science and technology, tactics, command and control, logistics, lethality, geopolitics, demographics, economics, or other relevant factors. Some attempts along these lines are reported in **Refs 7-19 through 7-32.** A list of some of the major wars and other significant events of the period 1600-1985 would illustrate the dramatic changes in science and technology over this period, and may be useful in relating combat trends to other historical events. Figure 7-1 shows the trend in one measure of weapon lethality over time (the index of lethality is explained in **Refs 7-22 and 7-30**).

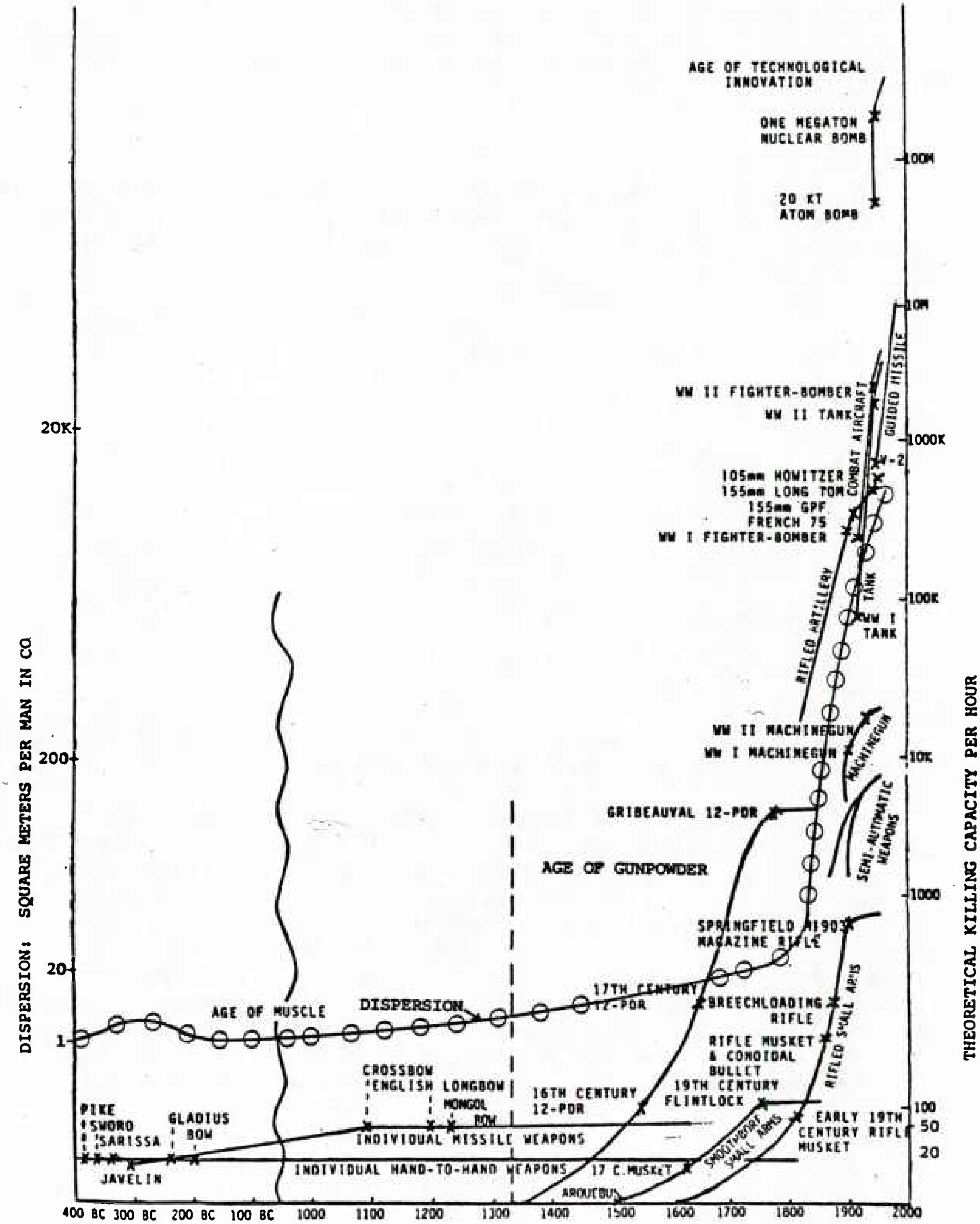


Figure 7-1. Increase of Weapon Lethality and Dispersion over History

7-5. CONCLUDING OBSERVATIONS ON OTHER ANALYSES

a. Some work by other investigators, cited in this chapter, suggests that certain long-term trends may be detectable in the data. In addition, information on rates of advance and on the influence air support has on land combat operations would be of value to wargamers and other military operations analysts. Although work on these topics is beyond the scope of the effort described in this paper, we plan to address them in future analyses.

b. The analysis of rates of advance will be greatly aided by availability of the CDES contract results regarding accurate battle durations.

c. Resolving the WWII anomaly will help identify those WWII battles that will provide the most trustworthy basis for examining the effects of tactical air support on the outcomes of land combat operations.

d. Future studies of long-term trends can profitably begin with a detailed re-examination of the long-term trends suggested by earlier works.

CHAPTER 8

CONCLUDING FINDINGS AND OBSERVATIONS

8-1. GENERAL. This paper documents the progress made on the CHASE Study during the period August 1984 - June 1985. During that period all tabular data in the HERO data base was reduced to machine readable form and subjected to a preliminary analysis. The appropriate next steps were also outlined. Efforts were made to adhere consistently to high standards of scientific practice.

8-2. KEY FINDINGS

a. Essential Elements of Analysis (EEA). The research was guided by five EEAs, as provided by the Study Directive (Appendix B). Summaries of the state of development reached during the period covered by this paper are as follows:

(1) Can the factors that have historically been most closely associated with victory in battle be identified? Six variables were tested for close association with victory in battle. Of these, three (ADV, LOG(FER), and RESADV)) seem technically much more closely associated with victory than the others (LOG(CER), LOG(EPS), and LOG(FR)). The battle data from World War II seems to be anomalous in the sense that the relationship of victory in battle to ADV seems to be much weaker than for battles of earlier and later eras. The reasons for this anomaly are not yet well understood, but the leading hypothesis seems to be that the data for several World War II era battles are flawed.

(2) What long-term trends can be detected in historical combat data? The analysis of long-term trends was not emphasized during the period covered by this paper. However, it appears that there has been no long-term secular trend over the last 400 years in the proportion of battles won by the attacker.

(3) Can the historical influence of air support on the outcome of land battles be quantified? An analysis of the effects of air support was not within the scope of the effort covered by this paper.

(4) What can be said about the factors influencing rates of advance in land combat? An analysis of the factors influencing rates of advance was not considered fruitful during the period covered by this paper, because the battle duration data in the data base used were reported only to the nearest day, which is too coarse a time resolution to provide rate values suitable for analysis.

(5) What lessons were learned regarding the preparation of battle and engagement data bases for use in quantitative analyses? Lessons learned regarding the preparation of data bases will be reported separately.

8-3. OBSERVATIONS

a. Observations on Data Bases

(1) The HERO data base needs to be enhanced before analyzing it extensively. To satisfy the need for data base refinement, the CHASE Data Enhancement Study (CDES) contract was awarded to the Historical Evaluation and Research Organization (HERO) to revise and extend the data base. The results of the CDES contract were not available in time to include in this paper.

(2) The HERO data base of 601 battles provides more detailed and systematically tabulated information on more battles, especially recent battles, than any other currently available data base. As a result it often is better suited to quantitative analysis than other sources of information. The CDES contract results will substantially enhance its accuracy and utility.

(3) Other, less comprehensive data bases will usefully supplement information in the HERO data base, and can be used selectively to investigate the extent to which findings based on the HERO data generalize readily to other data bases.

b. Observations on Descriptive Statistics

(1) Descriptive statistics express succinctly the predominant characteristics of a mass of data, and provide insights that usefully supplement those obtained by a study of individual cases. However, a clear perception of cause and effect relationships usually requires more sophisticated techniques.

(2) The HERO data base is mainly representative of short, pitched land combat battles fought by organized division- and corps-sized military formations during the 19th and early 20th centuries in Europe or North America.

(3) The attacker won about 61 percent of the 601 battles recorded in the HERO data base. The probability of an attacker victory may have declined slightly from 1600 to about 1850-1900, and then risen from about 1850-1900 to the 1970's, but the evidence for this gradual secular change is too slight to be depended upon.

(4) Battle durations seem to be distributed approximately as Weibull or as lognormal random variables.

(5) Casualty fractions seem to be distributed approximately lognormally. The attacker's casualty fraction tends to be less than the defender's.

(6) The attacker's personnel force ratio seems to be distributed roughly as a lognormal random variable. The attacker outnumbers the defender by a 3-to-1 margin in only about one-sixth of the battles. Victory seems to depend somewhat on force ratio, but not in a particularly reliable way. A 3-to-1 force ratio is neither necessary nor sufficient to assure victory in battle.

(7) The defender's personnel casualty exchange ratio is distributed approximately as a lognormal random variable. Since its median value is close to unity, the attacker's personnel casualties outnumber the defender's in about half the battles.

(8) The defender's personnel fractional exchange ratio seems to be distributed roughly as a lognormal random variable. It is less than unity in about two-thirds of the battles.

c. Observations on Factors Associated with Victory

(1) The variables ADV, LOG(FER), RESADV, LOG(CER), LOG(EPS) and LOG(FR) were compared with regard to the closeness of their association with victory in non-WWII battles, and were found to rank (from more closely associated to least) in the order listed. ADV, LOG(FER), and RESADV are nearly equally closely associated with victory in battle. The association between LOG(FR) and victory is not as close as any of the other five variables examined. Force ratio is an unsatisfactory and inadequate predictor of victory in battle. Both advantage and fractional exchange ratio are much more closely related to victory than is the force ratio. Consequently, either advantage or fractional exchange ratio should be used as a figure of merit for comparing force structures, contingency plans, equipment options, and tactics.

(2) Some of the battles in the HERO data base are anomalous, in the sense that their outcomes differ sharply from what is anticipated on the basis of the association of victory with ADV. A high proportion of the anomalous battles took place in the post-1940 era, even though most of these battles are not anomalous. In particular, the Italian, Northwest Europe, Okinawan, and 1973 October War (Golan Front) campaigns all seem to have relatively high incidences of anomalous battles. But the North African, Tarawa, Iwo Jima, Eastern Front, 1967 Six-Day and 1968 Arab-Israeli Wars, and 1973 October War (Suez Front) campaigns all seem to have about the same incidence of anomalous battles as do the battles of the pre-WWII era. Various hypotheses as to the cause of this WWII anomaly were presented and discussed. While the issue has not been definitively resolved, internal and circumstantial evidence suggests that the WWII anomaly could well be due to flaws in the data for some of the post-1940 battles. The planned independent review and reassessment of the data on the anomalous battles will provide valuable data on which to base a determination of the extent to which the WWII anomaly is a reflection of flawed data, or is due to some previously unanticipated phenomenon.

(3) Despite the WWII anomaly issue, ADV (or, alternatively, LOG(FER)) has been shown both theoretically and empirically to be substantially more accurate than other figures of merit for comparing the "military worth" of alternative materiel, organizations, and tactics.

d. Observations on the Analysis of Redundancy

(1) There is a high degree of redundancy among some of the items in the data base. The analysis of this redundancy, and the development of measures to deal correctly and effectively with it, need further investigation.

(2) The problem of determining what underlying or "basic" factors undergird a given set of observations has vexed scientists and philosophers for thousands of years. Factor analysis is currently reputed to be one of the most frequently used techniques, and we have applied it to 29 variables from the HERO data base. The results indicate that the original 29 variables can be replaced in future analyses by 8 new factors that are uncorrelated (i.e., not redundant) with practically no loss of information (in the technical sense).

(3) Nevertheless, the 8 new factors produced by the principal factors method may not be as intuitively clear as the original 29 were. Also, the present analysis intermingles original HERO variables from Tables 4 and 6, although there are good reasons to keep them separate. Moreover, we have applied factor analysis methods to categorical data even though the method technically requires the data to be continuous. Accordingly, the present exploratory effort at redundancy analysis must be used with caution and future analyses should consider alternative approaches.

e. Observations on Hypothesis Testing

(1) We have presented a test of a breakpoint hypothesis to illustrate the potential of hypothesis testing as a method for using combat data to study wargaming issues. This work may also serve as an instructive example for future efforts at hypothesis testing.

(2) The particular breakpoint hypothesis considered was shown to be false. This result casts serious doubt on the validity of the break curves conventionally used in wargames.

(3) Devising a satisfactory theory of victory in tactical operations was not within the scope of the effort reported in this paper.

f. Observations on Other Analyses. Some earlier work by other investigators suggests that certain long-term trends may be detectable in the data. In addition, information on rates of advance and on the influence of air support upon land combat operations would be of value in wargames and other military operations analyses. Although work on these topics is beyond the scope of the effort described in this paper, we plan to address them in future analyses. Future studies of long-term trends can profitably begin with a detailed reexamination of the long-term trends suggested by earlier works. The analysis of rates of advance will be greatly aided by the accurate battle durations to be made available by the CDES contract. Resolving the WWII anomaly will help identify which of the WWII battles will provide the most trustworthy basis for examining the effects of tactical air support on the outcomes of land combat operations.

8-4. CONCLUDING REMARKS. The findings and observations described in this paper were reached in the relatively short period of time between August 1984 and June 1985. Future efforts can profitably expand on and refine these results by more precise and detailed analyses of the data. Integration of the findings into a unified theoretical structure, while a desirable long-range goal, may be premature until empirical "laws" succinctly summarizing large areas of experience are formulated. Future work on CHASE should bear in mind the need for such "laws of combat," and seek to express them whenever the available data justify their formulation.

APPENDIX A

STUDY CONTRIBUTORS

1. STUDY TEAM

a. Study Director

Dr. Robert L. Helmbold, Requirements and Resources Directorate

b. Team Member

Dr. Aqeel Khan, Research and Analysis Support Directorate

2. PRODUCT REVIEW BOARD

Dr. Alan Johnsrud, Chairman
Dr. Jerome Bracken
LTC John R. Cary
MAJ David J. Fowler
Ms. Julia A. Fuller
Ms. Vera W. Hayes
Mr. Bradley Hill
Mr. Robert McQuie

APPENDIX B

STUDY DIRECTIVE

DEPARTMENT OF THE ARMY
US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

REPLY TO
ATTENTION OF:

CSCA-ZA

29 AUG 1984

MEMORANDUM FOR ASSISTANT DIRECTOR, STRATEGY, CONCEPTS AND PLANS

SUBJECT: Combat History Analysis Study Effort (CHASE)

1. PURPOSE OF STUDY DIRECTIVE. This directive provides tasking and guidance for the conduct of the Combat History Analysis Study Effort, which will perform an analysis of historical data on battles and engagements.

2. BACKGROUND. The Historical Evaluation and Research Organization (HERO) has recently presented a new data base of information on historical battles. This compilation is extensive and detailed for individual battles. In its present form, however, it is not directly useable in military operations research, concept formulation, war games, and studies requiring summary quantitative relationships applicable throughout a broad range of engagement situations.

3. STUDY SPONSOR AND SPONSOR'S STUDY DIRECTOR. US Army Concepts Analysis Agency (CAA) will sponsor this study. The Sponsor's Study Director will be Dr. Robert L. Helmbold of the Strategy, Concepts and Plans Directorate.

4. STUDY AGENCY. CAA's Strategy, Concepts and Plans Directorate will conduct this study. Augmentation and assistance will be provided as outlined in Paragraph 6 of this study directive.

5. TERMS OF REFERENCE

a. Scope

(1) Reduce all or a significant portion of the HERO data to machine-readable form for analysis.

(2) Summarize the mass of data and present it for use in military operations research, concept formulation, war gaming, and studies and analyses.

2 9 AUG 1984

CSCA-ZA
SUBJECT: Combat History Analysis Study Effort (CHASE)

(3) Seek trends and interrelations present but hidden in the data.

(4) Test selected hypotheses against the data.

b. Objective. Search for historically-based quantitative results for use in military operations research, concept formulation, war gaming, and studies and analyses.

c. Timeframe. Not applicable.

d. Assumptions

(1) Historical data can be treated as a statistical sample of possible outcomes. However, because there may be gross errors and biases in these data, robust statistical methods may be appropriate and confidence levels may have to be taken higher than usual to justify rejection of null hypotheses.

(2) Formulae are not to be complicated without good evidence.

(3) Trends and relationships that have persisted for a long period of time can be extrapolated into the forseeable future.

e. Essential Elements of Analysis (EEA)

(1) Can the factors that have historically been most closely associated with victory in battle be identified?

(2) What long-term trends can be detected in historical combat data?

(3) Can the historical influence of air support on the outcome of land battles be quantified?

(4) What can be said about the factors influencing rates of advance in land combat?

(5) What lessons were learned regarding the preparation of battle and engagement data bases for use in quantitative analyses?

CSCA-ZA 2 9 AUG 1984
SUBJECT: Combat History Analysis Study Effort (CHASE)

f. Environmental and Threat Guidance. No environmental consequences are envisioned; however, the study agency is required to surface and address any environmental considerations that develop in the course of the study effort.

g. Estimated Cost Savings or Other Benefits. Army studies and analyses often need summary quantitative relationships applicable throughout a broad range of combat situations. It would be costly and inefficient to have each study perform its own analysis of the historical data. Making the results of this study available will help avoid unnecessary duplication of analysis effort.

6. RESPONSIBILITIES. CAA's Strategy, Concepts and Plans Directorate will conduct the study. Assistance in keypunching data, developing or selecting appropriate statistical methods, and in performing statistical computations will be provided by CAA's Computer Support Directorate.

7. LITERATURE SEARCH. The principal source of historical combat data will be the Historical Evaluation and Research Office (HERO) report "Analysis of Factors That Have Influenced Outcomes of Battles and Wars; A Data Base of Battles and Engagements," Vols. I-VI, June 1983.

8. REFERENCES

a. AR 5-5, Army Studies and Analyses.

b. AR 10-38, Organization and Functions, United States Army Concepts Analysis Agency.

c. DA Pam 5-5, Guidance for Study Sponsors, Sponsor's Study Directors, Study Advisory Groups, and Contracting Officer Representatives.

d. DOD Directive 5010.22, The Management and Conduct of Studies and Analyses.

e. CAA Memorandum 5-1, Study Planning and Management.

f. CAA Memorandum 5-2, Quality Control of Agency Publications.

g. CAA Memorandum 310-3, Distribution of CAA Publications.

h. CAA Action Officer's Guide to Publication Services, April 1984 (CSCA-MSM-W).

CSCA-ZA 29 AUG 1984
SUBJECT: Combat History Analysis Study Effort (CHASE)

i. CAA Graphic Arts Policy and Procedure Guide, April 1984 (CSCA-MSM-G).

j. CAA Memorandum 310-6, Standards for CAA Briefings.

k. CAA Study Director's Guide, July 1983 (CSCA-MSM-O).

9. ADMINISTRATION

a. Resource costs (funds, manpower, computer time, TDY, and administrative support) will be borne by CAA.

b. Administrative support such as clerical, office space, office equipment, etc., will be furnished by CAA.

c. It is anticipated that no more than 15 Professional Staff Months (PSM) will be expended.

d. It is anticipated that no more than 200 hours of computer time will be needed for statistical and other computerized analysis.

e. Milestone schedule

(1)	Study directive approval (Dir, CAA)	Aug 84
(2)	Select and analyse exploratory data base	Aug 84 - Feb 85
(3)	Select and analyse confirmatory data base	Dec 84 - Sep 85
(4)	Draft final report	Jul 85 - Nov 85
(5)	Draft final to PRB	Nov 85
(6)	Revise final report	Nov 85
(7)	Publish final report	Nov 85

Quarterly progress memoranda reports should be submitted to Director, CAA through Assistant Director, SPP.

f. CAA Will prepare and submit DD Form 1498 and final study documents to DTIC.

29 AUG 1984

CSCA-ZA
SUBJECT: Combat History Analysis Study Effort (CHASE)

g. Final documentation should be in the form of a CAA Technical Paper that describes the study's findings and documents their technical basis.

h. A statement of lessons learned, including any appropriate recommendations for continuing or follow-on historical analysis efforts, will be provided to Dir, CAA in the form of an internal CAA memorandum.

signed

DAVID C. HARDISON
Director

CF
Assistant Director, CP

APPENDIX C

REFERENCES

1-1. "Analysis of Factors That Have Influenced the Outcomes of Battles and Wars: A Data Base of Battles and Engagements," Vols I-VI, CAA-SR-84-6, September 1984

1-2. Platt, John R., "Strongs Inference," Science, Volume 146 (1964), No. 3642 (16 October), pp. 347-353

1-3. Chamberlin, T. C., "The Method of Multiple Working Hypotheses," Science, Volume 148 (1965), 7 May, pp. 754-759

1-4. Article on "William of Ockham," Encyclopedia Brittanica, 1974

1-5. Richardson, Lewis Fry, Statistics of Deadly Quarrels, Boxwood Press, Pacific Grove, CA, 1960

1-6. Koopman, Bernard O., "An Operational Critique of Detection Laws," Operations Research, Volume 27 (1979), No. 1 (Jan-Feb), pp. 115-133

1-7. Helmbold, Robert L., Unpublished, 1970

1-8. Helmbold, Robert L., Unpublished, 1985

2-1. "Analysis of Factors That Have Influenced the Outcomes of Battles and Wars: A Data Base of Battles and Engagements," Vols I-VI, CAA-SR-84-6, September 1984

2-2. "Lanchester Parameters for Some Battles of the Last Two Hundred Years," CORG SP-122, Robert L. Helmbold, Combat Operations Research Group, HQ US Continental Army Command, 1961

2-3. "Historical Data and Lanchester's Theory of Combat," CORG SP-128 Robert L. Helmbold, Combat Operations Research Group, HQ US Continental Army Command, 1 July 1961

2-4. "Historical Data and Lanchester's Theory of Combat: Part II," CORG SP-190, Robert L. Helmbold, Combat Operations Research Group, HQ US Army Combat Developments Command, August 1964

2-5. Computer file available through Dr. Robert L. Helmbold

2-6. Militar-Historisches Kriegs-Lexicon (1618-1905), C. W. Stern, Vienna & Leipzig, 1908

2-7. "Lanchester as Force in History: An Analysis of Land Battles of the Years (1618-1905)," Daniel Willard, RAC-TP-74, Research Analysis Corporation, November 1962

2-8. "The Use of Lanchester-Type Equations in the Analysis of Past Military Engagements," William A. Schmieman, Doctoral Thesis, Georgia Institute of Technology, August 1967

2-9. Numbers, Predictions, and War: Using History to Predict the Outcome of Battles, Trevor N. Dupuy, Robbs-Merrill Co., Indianapolis, IN, 1979

2-10. Numbers and Losses in the Civil War in America: 1961-1865, Thomas L. Livermore, Houghton-Mifflin, Boston, MA, 1901. (Reprinted by Kraus Reprint Co., Millwood, NY, 1977)

2-11. Napoleon: A History of the Art of War, from Lutzen to Waterloo, With a Detailed Account of the Napoleonic War, Theodore A. Dodge, (4 Vols), Boston and New York, 1904-1907. (Reprinted by AMS Press, Inc., New York 1970)

2-12. Die Zahl in Kriege: Statistiche Daten aus der Neuren Kriegeschicht in Graphischer Darstellung," Otto Berndt, G. Fregtag & Berndt, Vienna, 1897

3-1. Statistics of Deadly Quarrels, Lewis F. Richardson, Boxwood Press, Pacific Grove, CA, 1960

3-2. "A Statistical Model for the Duration of Wars and Strikes," William Horvath, Behavioral Science, 13 (Jan 1968), pp. 18-28

3-3. "The Lognormal Distribution of the Duration of Strikes," R. J. Lawrence, J. R. Statist, Soc. A, 147 (1984), Part 3, pp. 464-483

3-4. "Stochastic Models for the Duration and Magnitude of a 'Deadly Quarrel'," Herbert K. Weiss, Operations Research, 11 (1963), pp. 101-121

3-5. Statistical Distributions: A Handbook for Students and Practitioners, N. A. J. Hastings and J. B. Peacock, Butterworths, London, 1975

3-6. Handbook of Tables for Probability and Statistics, William H. Beyer, Second Edition, CRC Press, Boca Raton, FL, Fifth Printing, 1981

3-7. Probability and Its Engineering Uses, Thorton C. Fry, D. Van Nostrand Co., Princeton, NJ, 1928

3-8. Statistics Manual, Edwin L. Crow, F. A. Davis, and M. W. Maxfield, NAVORD REPORT 3369 - NOTS 948. Republished 1960 by Dover Publications, New York

3-9. The Theory of Probability, B. V. Gnedenko, Chelsea Publishing Co., New York, 1963

3-10. On War, Carl von Clausewitz, Edited and Translated by Michael Howard and Peter Paret, Princeton University Press, Princeton, NJ, 1976

3-11. "Technology and Tactics in Defense of Europe," General William E. Depuy, US Army, Retired, Army Magazine, April 1979, pp. 14-23

3-12. "The Influence of the Numerical Strength of Engaged Sides on Their Losses," (in Russian), M. Osipov. A five-part article published in Voenniyy Sbornik, 1915, as follows, Part 1, No. 6 (Jan), pp. 31-40; Part 2, No. 7 (Jul), pp. 25-361; Part 3, No. 8 (Aug), pp. 31-40; Part 4, No. 9(Sep), pp. 25-37; Part 5, No. 10 (Oct), pp. 93-96

3-13. "Historical Data and Lanchester's Theory of Combat," Robert L. Helmbold, SP-128, Combat Operations Research Group, HQ US Army Continental Army Command, 1 July 1961

3-14. "Historical Data and Lanchester's Theory of Combat - Part II," Robert L. Helmbold, SP-190, Combat Operations Research Group, HQ US Army Combat developments Command, August 1964

3-15. "The Use of Lanchester-Type Equations in the Analysis of Past Military Engagements," William A. Schmieman, Doctoral Thesis, Georgia Institute of Technology, August 1967

3-16. "A Lanchester Fit to Selected Land Battles," Lulejian & Associates, Contract F44620-70-C-0056, February 1971

3-17. "Probability of Victory in Land Combat as Related to Force Ratio," Rand Paper P-4199, Robert L. Helmbold, The Rand Corporation, Santa Monica, CA October 1969

4-1. Aldrich, John H. and Forrest D. Nelson, Linear Probability, Logit, and Probit Models, Sage Publications, Beverly Hills, CA, 1984

4-2. Cox, D. R., The Analysis of Binary Data, Chapman & Hall, London, 1970

4-3. Daganzo, Carlos, Multinomial Probit: The Theory and Its Application to Demand Forecasting, Academic Press, New York, NY, 1979

4-4. Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," CORG-SP-128, Combat Operations Research Group, HQ, US Army Continental Army Command, Fort Monroe, VA, 1 July 1961

4-5. Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat: Part II," CORG-SP-190, Combat Operations Research Group, HQ, US Army Combat Developments Command, Fort Belvoir, VA, August 1964

4-6. Helmbold, Robert L., et al., "Integrated Warfare--Representing the Decision Process Within Computer Simulations of Combat," CAA-TP-85-1, US Army Concepts Analysis Agency, Bethesda, MD, January 1985

4-7. Clausewitz, Carl Von, On War, edited and translated by Michael Howard and Peter Paret, Princeton University Press, Princeton, NJ, 1976

4-8. Crow, Edwin L., Frances A. Davis and Margaret W. Maxfield, Statistics Manual, NAVORD Report 3369-NOTS 948, Naval Ordnance Test Station, China Lake, CA. 1955 (Republished by Dover Publications, Inc., New York, 1960)

4-9. Beyer, William H., Handbook of Tables for Probability and Statistics, CRC Press, Inc., Boca Raton, Florida, 1968 (Fifth Printing, 1981)

4-10. Breiman, Leo and Jerome H. Friedman, "Estimating Optimal Transformation for Multiple Regression and Correlation," Journal of the American Statistical Association, Vol. 80 (1985), No. 391 (September), pp. 580-619

5-1. Nie, Norman H., et al., SPSS Statistical Package for the Social Sciences, Second Edition, McGraw Hill Book Company, New York, NY, 1975

5-2. Morrison, Donald F., Multivariate Statistical Methods, McGraw Hill Book Company, New York, NY, 1967

5-3. Snedecor, George W., and William G. Cochran, Statistical Methods, Fifth Edition, The Iowa State University Press, Ames, Iowa, 1961

5-4. Harman, Harry H., Modern Factor Analysis, Second Edition, Revised, The University of Chicago Press, Chicago, IL, 1968

5-5. Hald, A., Statistical Theory with Engineering Applications, John Wiley & Sons, New York, NY, 1960

5-6. Wu, Teh-Wei, Econometrics: An Introductory Analysis, University Park Press, Baltimore, MD, 1973

5-7. Jure Zupan, Clustering of Large Data Sets, Research Studies Press, A Division of John Wilely & Sons, Ltd., Chichester, England, 1982

5-8. Huber, Peter J., "Projection Pursuit," Annals of Statistics, Vol. 13(1985), No. 2(Jun), pp. 435-525

5-9. Brieman, Leo and Jerome H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," Journal of the American Statistical Association, Vol. 80(1985), No. 391(September), pp. 580-619

5-10. Rao, C. Radhakrishna, Linear Statistical Inference and Its Applications, John Wiley & Sons, Inc., New York, 1965

5-11. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Second Edition, John Wiley & Sons, Inc., New York, 1984

5-12. Draper, Norman R. and Harry Smith, Applied Regression Analysis, Second Edition, John Wiley & Sons, Inc., New York, 1980

5-13. Kendall, Maurice G., Multivariate Analysis, Charles Griffin & Co., London, 1975

6-1. Helmbold, Robert L., "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," Rand Report R-772-PR, The Rand Corporation, Santa Monica, California, June 1976

6-2. "Maneuver Control," Department of the Army Field Manual, FM 105-5, April 1964

6-3. Clark, Dorothy K. "Casualties as a Measure of the Loss of Combat Effectiveness of an Infantry Battalion," ORO-T-289, Operations Research Office, The Johns Hopkins University, August 1954

6-4. Weiss, Herbert K., "Stochastic Models for the Duration and Magnitude of a 'Deadly Quarrel'," Operations Research, Vol. 11 (1963), pp. 101-121

6-5. Horvath, William J., "A Statistical Model for the Duration of Wars and Strikes," Behavioral Science, Vol. 13, January 1968, pp. 18-28

6-6. Weiss, Herbert K., "Combat Models and Historical Data: The US Civil War," Operations Research, Vol. 14 (1966), pp. 759-790

6-7. Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," CORG-SP-128, Combat Operations Research Group, HQ, US Army Continental Army Command, Fort Monroe, Virginia, 1 July 1961

6-8. Richardson, Lewis F., "Statistics of Deadly Quarrels," Boxwood Press, Pacific Grove, California, 1960

7-1. Willoughby, Charles A., Maneuver in War, The Military Publishing Co., Harrisburg, PA, 1939

7-2. Parsons, COL Norman W., "Military Unit Rates of Advance," Combat Operations Research Group, Ft Monroe, VA, 1954

7-3. Hulse, BG A. D., "Movement," Combat Operations Research Group, Ft Monroe, VA, 1954

7-4. Andrews, Marshall, "Rates of Advance in Land Attack Against Unprepared Forces," ORO-TP-10, Operations Research Office, Bethesda, MD, August 1960

7-5. "Analysis of Fundamental Relations in Army Mobility Systems: Vol IV, Determination of the Relation Between Mobility and Combat Effectiveness by Historical Analysis," RAC-TP-268, Research Analysis Corporation, McLean, VA, February 1968

7-6. Wainstein, L., "An Examination of the Parsons and Hulse Papers of Rates of Advance," Institute for Defense Analyses, Alexandria, VA, December 1973

7-7. Goad, Rex, "Prediction of Opposed Movement and Casualty Rates," SHAPE Technical Centre, Professional Paper, STC-PP-75, July 1973

7-8. Goad, Rex, "Predictive Equations for Opposed Movement and Casualty Rates for Land Forces," SHAPE Technical Centre, Professional Paper, STC-PP-90, August 1974

7-9. "A Survey of 'Quick Wins' in Modern War," AD-A025 893, Historical Evaluation and Research Organization, Dunn Loring, VA, October 1975

7-10. Goad, Rex, "The Modeling of Movement in Tactical Games," Professional Paper, STC-PP-158, SHAPE Technical Centre, The Hague, April 1979

7-11. Hardison, David C., "Movement Rates--CEM IV," CAA Internal Memorandum, 1984

7-12. Savkin, Vasillii E., Rates of Advance, (in Russian), Military Publishing House, Ministry of Defense, Moscow, 1965

7-13. Delbruck, Hans, Geschicte der Kriegskunst, (6 volumes.), Berlin, 1907-1929. (Republished as History of the Art of War Within the Framework of Political Philosophy, Greenwood Press, Westport, Conn, Volume 1 (1975), Volume 2 (1980))

7-14. Richardson, Lewis Fry, Statistics of Deadly Quarrels, Boxwood Press, Pacific Grove, CAL, 1960

7-15. Smith, E. G., and G. N. Donovan, "Trends in Warfare," Memorandum No. F-6, British Army Operational Research Group, April 1955

7-16. Naroll, Raoul, Vern L. Bullough and Frada Narroll, Military Deterrence in History, State University of New York Press, 1974

7-17. Voevodsky, John, "Crisis Waves: The Growth and Decline of War-Related Events," Journal of Psychology, Volume 80 (1972), pp 287-308

7-18. Singer, J. David, and Melvin Small, The Wages of War 1816-1965: A Statistical Handbook, John Wiley & Sons, Inc., New York, 1972

7-19. Sorokin, Pitirim A., "Social and Cultural Dynamics" (4 volumes), 1937-1941 (reissued 1962)

7-20. Bodart, Gaston, Military-Historisches Kriegslexican (1618-1905), C. W. Stern, Vienna & Leipzig, 1908

7-21. Denton, Frank H., "Factors in International System Violence," Rand Paper P-2416, The Rand Corporation, Santa Monica, CAL, October 1969

7-22. Dupuy, Trevor N., The Evolution of Weapons and Warfare, Bobbs-Merrill Co., Indianapolis, Ind, 1980

7-23. Haustein, Heinz-Dieter, and Erich Neuwirth, "Long Waves in World Industrial Production, Energy Consumption, Innovation, Inventions, and Patents and Their Identification by Spectral Analysis," Technological Forecasting and Social Change, Volume 22 (1982), No. 1 (Sep), pp 53-89

7-24. Thompson, William R., and L. Gary Zuk, "War, Inflation, and the Kondratieff Long Wave," Journal of Conflict Resolution, Volume 26 (1982), No. 4 (Dec), pp 621-644

7-25. Low, William, "Discoveries, Innovations, and Business Cycles," Technological Forecasting and Social Change, Volume 26 (1984), No. 4 (Dec), pp 355-373.

7-26. Craig, Paul P. and Kenneth E. F. Watt, "The Kondratieff Cycle and War: How Close is the Connection?," The Futurist, April 1985, pp 25-27

7-27. Schrodt, Philip A., "Microcomputers in the Study of Politics: Predicting Wars With the Richardson Arms-Race Model," BYTE, July 1982, pp 108-134

7-28. Ward, Michael D., "Simulating the Arms Race," BYTE, October 1985, pp 213-222

7-29. Hibbs, Douglas A., Jr., Mass Political Violence: A Cross-National Causal Analysis, John Wiley & Sons, New York, 1973

7-30. Hughes, B. P., Firepower: Weapons Effectiveness on the Battlefield, 1630-1850, Charles Scribner's Sons, New York, 1974

7-31. Rosenau, James N. (Ed.), In Search of Global Patterns, The Free Press, New York, 1976

7-32. Wright, Quincy, A Study of War, (2 volumes), University of Chicago Press, 1942, (1 volume) Second Edition, 1965, Second Impression, 1967

J-1. Kendall, Maurice G. and Alan Stuart, The Advanced Theory of Statistics (three volumes), Charles Griffin & Co., Ltd., London (1969-1968)

J-2. Cramer, Harald, Mathematical Methods of Statistics, Princeton University Press, Princeton, NJ, 1946

J-3. Cox, D. R., The Analysis of Binary Data, Chapman & Hall, London, 1970

J-4. Aldrich, John H. and Forrest D. Nelson, Linear Probability, Logit, and Probit Models, Sage Publications, Beverly Hills, CA, 1984

J-5. Daganzo, Carlos, Multinomial Probit: The Theory and Its Application to Demand Forecasting, Academic Press, New York, NY, 1979

APPENDIX D

BIBLIOGRAPHY

AUTHOR BIBLIOGRAPHY OF WORKS ON THE THEORY OF COMBAT

- A -

Adkins, Richard C., "Analysis of Unit Breakpoints in Land Combat," Student Thesis, US Naval Postgraduate School, Monterey, CA, March 1975.

Aldrich, John H. and Forrest D. Nelson, "Linear Probability, Logit, and Probit Models," Sage Publications, Beverly Hills, CA 1984.

Alker, H. R., Jr., "Research Paradigms and Mathematical Politics," To be published in "Proceedings of the IPSA Roundtable," Edited by Rudolf Wildenmann, et al, 1973.

Anderson, T. W., "An Introduction to Multivariate Statistical Analysis," Second Edition, John Wiley & Sons, Inc., New York, 1984.

Andrews, Marshall, "Rates of Advance in Land Attack Against Unprepared Forces," ORO-TP-10, Operations Research Office, Bethesda, MD, August 1960.

Anonymous, "Civil War: Part II; Now History, The Battles," Life Magazine, 20 January 1961.

Anonymous, "Historical Statistics," Statistisk Sentralbyra, Oslo, Norway, 1968.

Anonymous, "Bibliography: World War II Tactical Air Interdiction," USAF Studies and Analysis, November 1970.

Anonymous, "Staff Officer's Field Manual Organizational, Technical and Logistial Data (Unclassified Data)," FM 101-10-1, HQ Department of the Army, July 1971.

Anonymous (Atkeson, MG Edward B.), "In Pursuit of the Essence of War," The Army Magazine, January 1984, pp. 18-24.

Anonymous, "Analysis of Fundamental Relations in Army Mobility Systems," Research Analysis Corporation, Date unknown.

Armstrong, J. Scott, "Advocacy and Objectivity in Science," Management Science, Vol. 25(1979), No. 5(May), pp. 423-428.

Army, Department of the, "Maneuver Control," Department of the Army Field Manual, FM 105-5, April 1964.

- B -

Bacon, Francis, "Novum Organum," 1620.

Banks, Arthur S., et al, "Cross-Polity Time Series Data," The MIT Press, Cambridge, MA 1971.

Barber, Bernard, "The Resistance of Scientists to Scientific Discovery," Science, v. 134(1961), pp. 596-602.

Barringer, Richard E., "War: Patterns of Conflict," MIT Press, Cambridge, MA, 1972.

Barringer, Richard E., "War: Technical Manual," MIT Press, Cambridge, MA, 1972.

Beebe, Gilbert W., and Michael E. DeBakey, "Battle Casualties," Charles C. Thomas, Publisher, Springfield, IL, 1952.

Beer, Francis A., "The Political Economy of Alliances: Benefits, Costs, and Institutions in NATO," Volume 1 of the International Studies Series, Sage Publications, Beverly Hills, CA, 1972.

Bennett, Jay M., and John A. Flueck, "An Evaluation of Major League Baseball Offensive Performance Models," The American Statistician, v. 37(1983), No. 1(Feb), pp. 76-82.

Berelson, Bernard, and Gary A. Steiner, "Human Behavior: An Inventory of Scientific Findings," Harcourt, Brace & World, Inc., New York, 1964.

Berndt, Otto, "Die Zahl im Krieg: Statistische Daten aus der Neuren Kriegeschicht in Graphischer Darstellung," G. Freytag & Berndt, Vienna, 1897.

Best, Robert J., "Analysis of Personnel Casualties in the 25-th Infantry Division, 26-31 July 1950," ORO-T-22(FEC), Operations Research Office, 14 April 1952.

Best, Robert J., "A Study of Battle Casualties Among Equivalent Opposing Forces (Korea, September 1950)," Tehnical Memo ORO-T-23(FEC), Operations Research Office, 1952.

Beyer, William H., "Handbook of Tables for Probability and Statistics," Second Edition, CRC Press, Boca Raton, FL, Fifth Printing, 1981.

Bodart, Gaston, "Mitary-Historisches Kriegs-Lexicon (1618-1905)," C. W. Stern, Vienna & Leipzig, 1908.

Bodart, Gaston, and Vernon L. Kellogg, "Losses of Life in Modern Wars," Clarendon Press, Oxford, 1916.

Boulding, Kenneth E., "Conflict and Defense: A General Theory," Harper, 1962.

Box, G. E. P., "The Use and Abuse of Regression," Technometrics, v. 8(1966), pp. 625-629.

Brieman, Leo and Jerome H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," Journal of the American Statistical Association, Vol 80(1985), No 391(September), pp 580-619.

Buell, J. D., H. H. Kagiwada, and R. E. Kalaba, "Quasilinearization and Inverse Problems for Lanchester Equations of Conflict," Operations Research, v. 16(1968), no. 2(Mar-Apr), pp. 437-442.

Bueno do Mesquita, Bruce, "The War Trap," Yale University, New Haven, CT, 1981.

Busse, James J., "An Attempt to Verify Lanchester's Equations," 1971. (In the book "Developments in Operations Research," by Benjamin Avi-Itzhak (Ed.), Gordon & Breach Science Publishers, 1971.)

- C -

Calahan, H. A., "What Makes a War End," Vanguard Press, New York, 1944.

Campbell, Donald T., and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research," Rand-McNally, Chicago, 1966.

Carroll, Bernice A., "How Wars End: An Analysis of Some Current Hypotheses," Journal of Peace Research, v. 4(1969), pp. 295-323.

Chamberlin, T. C., "The Method of Multiple Working Hypotheses," Science, v. 148(1965), pp. 754-759.

Clark, Dorothy K., "Casualties as a Measure of the Loss of Combat Effectiveness of an Infantry Battalion," ORO-T-289, Operations Research Office, August 1954.

Clausewitz, Carl von, "The Most Important Principles of War for the Conduct of War," Stackpole Books. (Reprinted December 1983 by US Army War College, Carlisle, PA.)

Clausewitz, Karl, "On War," Edited and Translated by Michael Howard and Peter Paret, Princeton University Press, Princeton, NJ, 1976.

Cochran, William G., "The Planning of Observational Studies of Human Populations," Journal of the Royal Statistical Society, Series A, v. 132(1965), pp. 234-255.

Cochran, William G., and D. Rubin, "Controlling Bias in Observational Studies: A Review," Sankhya, v. A35(1973), pp. 417-446.

Couchri, Nazli, and Robert C. North, "Nations in Conflict: Domestic Growth and International Violence," Freeman, San Francisco, CA, 1975.

Cox, D. R., "The Analysis of Binary Data," Chapman & Hall, London, 1970.

Craig, Paul P. and Kenneth E. F. Watt, "The Kondratieff Cycle and War: How Close is the Connection?", The Futurist, April 1985, pp. 25-27.

Cramer, Harald, "Mathematical Methods of Statistics," Princeton University Press, Princeton, NJ, 1946.

Crow, Edwin L., F. A. Davis and M. W. Maxfield, "Statistics Manual," NAVORD REPORT 3369 - NOTS 948. Republished 1960 by Dover Publications, New York.

- D -

Daganzo, Carlos, "Multinomial Probit: The Theory and Its Application to Demand Forecasting," Academic Press, New York, 1979.

Dahm, Ulrich, "Eine Theorie Zur Beschreibung Der 'Schlact im Atlantik'," Angewandte Informatik, v. 2(1979), pp. 73-79.

Daly, F., "The UK Lanchester Study Group," Omega Int. J. of Mgmt. Sci, Vol. 13(1985), No. 2, pp. 131-133.

Daniel, Cuthbert, and Fred S. Wood, "Fitting Equations to Data: Computer Analysis of Multifactor Data," John Wiley & Sons, New York, 1980.

Delbruck, Hans, "Geschicte der Kriegskunst," (6 vols.), Berlin, 1909-1929. (Republished as "History of the Art of War Within the Framework of Political Philosophy," Greenwood Press, Westport, CT, Vol. 1(1975), Vol. 2(1980).)

Delbruck, Hans, "Numbers in History," University of London, London, 1913.

Denton, Frank H., "Factors in International System Violence," Rand Paper P-4216, The Rand Corporation, Santa Monica, CA, October 1969.

Depuy, Gen. William E., US Army (Ret), "Technology and Tactics in Defense of Europe," Army Magazine, April 1979, pp. 14-23.

Dodge, Theodore A., "Napoleon: A History of the Art of War," (4 vols), 1904-1907. (reprint) AMS Press, New York, 1970.

Dolansky, Ladislav, "Present State of the Lanchester Theory of Combat," Operations Research, v. 12(1964), no. 2(Mar-Apr): 344-358.

Draper, Norman R., and Harry Smith, "Applied Regression Analysis," John Wiley & Sons, New York, 1966. (Second Edition, 1980.)

Dumas, S., and K.O. Vedel-Petersen, "Losses of Life Caused by War," Oxford, 1923.

Dunn, L. P., "A Lanchester Fit to Selected Land Battles," Lulejian & Associates, Contract F44620-70-C-0056, February 1971.

Dupuy, R. E., and T. N. Dupuy, "Military Heritage of America," McGraw-Hill, 1956.

Dupuy, R. E., and T. N. Dupuy, "The Encyclopedia of Military History," Harper & Row, Publishers, New York, 1970.

Dupuy, Trevor N., and Grace P. Hayes, "Opposed Rates of Advance of Large Forces in Europe (ORALFORE)," AD 902 830, 28 Aug 1972.

Dupuy, Trevor N., "A Survey of 'Quick Wins' in Modern War," Historical Evaluation and Research Organization, Fairfax, VA, October 1975.

Dupuy, Trevor N., and Janice B. Fain, "Laws Governing Combat," National Defense, Nov-Dec 1975, 218-221.

Dupuy, Trevor N., et al, "A Study of Breakthrough Operations," AD A036 492, Oct 1976.

Dupuy, Trevor N., "The Current Implications of German Military Excellence," Strategic Review, v. 4(1976), no. 4 (Fall), pp. 87-94.

Dupuy, Trevor N., "Predicting the Element of Surprise in Combat," Army, May 1977, 44-51.

Dupuy, Trevor N., "Weapons Lethality and the Nuclear Threshold," Armed Forces Journal International, October 1978, pp. 24-33.

Dupuy, Trevor N., "Numbers, Predictions, and War: Using History to Predict the Outcome of Battles," Bobbs-Merrill Co., Indianapolis, IN, 1979.

Dupuy, Trevor N., "Perceptions of the Next War," Armed Forces Journal International, May 1980, pp. 49-65.

Dupuy, Trevor N., "The Evolution of Weapons and Warfare," Bobbs-Merrill Co., Indianapolis, IN, 1980.

Dupuy, Trevor N., and Paul Martell, "Great Battles on the Eastern Front: The Soviet-German War, 1941-1945," Bobbs-Merrill Co., Indianapolis, IN, 1982.

Dupuy, Trevor N., "Analysis of Factors That Have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements," (6 vols.), Historical Evaluation and Research Organization, Fairfax, VA, 1984.

Dupuy, Trevor N., "Criticisms of Combat Models Cite Unreliability of Results," Army, March 1985, pp. 16-18.

- E -

Eggenberger, David, "A Dictionary of Battles," Thomas Y. Crowell Co., New York, 1967.

Encyclopedia Brittanica, Article on "William of Ockham," 1974

Engel, Joseph H., "A Verification of Lanchester's Law," Operations Research, v. 2(1954), pp. 163-171.

Engel, Joseph H., "Combat Effectiveness of Allied and German Troops in the World War II Battle of Crete," OEG IRM-35, Operations Evaluation Group, 8 Feb 1963.

Engel, Joseph H., "Further Model Evaluation of the Battle of Iwo Jima," Presented at the ORSA/TIMS Meeting, 1984.

- F -

Fain, William W., Janice B. Fain, Leon Feldman, and Susan Simon, "Validation of Combat Models Against Historical Data," Professional Paper No. 27, Center for Naval Analyses, 1970.

Fain, Janice, "The Lanchester Equations and Historical Warfare: An Analysis of Sixty World War II Land Engagements," An UNCLASSIFIED paper contained in Proceedings of the 34-th Military Operations Research Symposium, December 3-5, 1974. (SECRET.) (Republished in History, Numbers, and War, v. 1(1977), Spring, pp. 34-52.)

Falls, C., "A Hundred Years of War, 1850-1950," Collier Books, New York, 1962.

Feldman, L., and S. Simon, "Study of Land/air Tradeoffs (SLAT)," Center for Naval Analyses," Contract N00014-68-0-0091, in 9 volumes, 1970. Vol. 7: "Historical Data Base." Vol. 8: "The Relation of the Tactical Warfare Simulation Program to Historical Warfare." Vol. 9: "Statistical Analysis of the Korean War."

Feller, William, "An Introduction to Probability Theory and Its Applications," (2 vols), John Wiley & Sons, New York, v. 1(1950), v. 2(1966).

Ferris, Wayne H., "The Power Capabilities of Nation-States: International Conflict and War," Lexington Books, Lexington, MA, 1973.

Fitch, Walter M. and Emmanuel Margoliash, "Construction of Phylogenetic Trees," Science v. 155(1967), no. 3760(20 Jan): 279-284.

Fleming, COL Norwood W., "An Analysis of Combat Attrition and Intensity of War (CATIWAR) (U)," CAA Technial Paper CAA-TP-80-4, US Army Concepts Analysis Agency, Bethesda, MD, April 1980. (SECRET-NOFORN)

Foch, Ferdinand, "The Principles of War," Chapman & Hall, London, 1918.

Fossum, E., "Factors Influencing the Occurrence of Military Coups d'Etat in Latin America," Journal of Peace Research, v. 3(1967), pp. 225-251.

Freeman, David A., "A Note on Screening Regression Equations," The American Statistician, Vol. 37, No. 2, May 1983, pp. 152-155.

Fry, Thornton C., "Probability and Its Engineering Uses," Van Nostrand Co., Princeton, NJ, 1928.

Fuchs, W., "Formeln Zur Macht," Deutsche Verlassanfalt, Stuttgart, 1965.

Fuller, John F. C., "The Foundations of the Science of War," Hutchinson & Co., London, 1925. (Reprinted December 1983 bu US Army War College, Carlisle, PA.)

Fuller, John F. C., "The Conduct of War, 1789-1961," Eyre and Spottiswoode, London, 1961.

- G -

Gardner, Martin, "Fads and Fallacies in the Name of Science," Dover Publications, New York,1957. (A reissue of "In the Name of Science," published il 1952).

Gillespie, John V. (Ed.), "Mathematical Systems In International Relations," Praeger Publishers, New York, 1977.

Gochman, Charles S., "Status, Conflict and War: The Major Powers, 1820-1970," University of Michigan, Ann Arbor, MI, 1975.

Gnedenko, B. V., "The Theory of Probability," Chelsea Publishing Co., New York, 1962.

Goad, Rex, "Prediction of Opposed Movement and Casualty Rates," SHAPE Technical Centre, Professional Paper, STC-PP-75, July 1973.

Goad, Rex, "Predictive Equations for Opposed Movement and Casualty Rates for Land Forces," SHAPE Technical Centre, Professional Paper, STC-PP-90, August 1974.

Goad, Rex, "The Modelling of Movement in Tactical Games," Professional Paper STC PP-158, Operations Research Division, SHAPE Technical Center, The Hague, April 1979.

- H -

Hald, A., "Statistical Theory With Engineering Applications", John Wiley & Sons, New York, NY, 1960.

Hammerman, G. M., "Conventional Attrition and Battle Termination Criteria: A Study of War Termination," AD A148 773, 01 Aug 1982.

Hammerton, Dr. M., "A Case of an Inappropriate Model," Nature, v. s203(1964), no. 4940(4 July): 64-64.

Harbison, Frederick H., Joan Maruhnic, and Jane R. Resnick, "Quantitative Analysis of Modernization and Development," Princeton University Press, Princeton, NJ, 1970.

Harbottle, Thomas B., "Dictionary of Battles," E. P. Dutton & Co., New York, 1905 (Republished 1966 by Gale Research Co., Detroit, MI).

Hardison, David C., "Movement Rates--CEM IV," CAA Internal Memorandum, 1984.

Harman, Harry H., "Factor Analysis", Second Edition, Revised, The University of Chicago Press, Chicago, IL, 1968.

Harville, David, "Experimental Randomization: Who Needs It?", The American Statistician, v. 29(1975), No. 1(Feb), pp. 27-31.

Hastings, N. A. J. and J. B. Peacock, "Statistical Distributions: A Handbook for Students and Practitioners," Butterworths, London, 1975.

Haustein, Heinz-Dieter, and Erich Neuwirth, "Long Waves in World Industrial Production, Energy Consumption, Innovation, Inventions, and Patents and Their Identification by Spectral Analysis," Technology Forecasting and Social Change, v. 22(1982), No. 1(Sep), pp. 53-89.

Heinl, Robert D., "Dictionary of Military and Naval Quotations," U. S. Naval Institute, Annapolis, MD, 1966.

Helmbold, Robert L., "Psychological Effects of Fire," CORG-M-73, Combat Operations Research Group, HQ US Continental Army Command, 1959.

Helmbold, Robert L., "Lanchester Parameters for Some Battles of the Last Two Hundred Years," CORG-SP-122, Combat Operations Research Group, HQ US Continental Army Command, 1961.

Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," Robert L. Helmbold, CORG-SP-128, Combat Operations Research Group, HQ US Continental Army Command, 1 July 1961.

Helmbold, Robert L., "Lanchester Equations, Historical Battles, and War Games(U)," Proceedings of the Eighth Military Operations Research Symposium, Pt. Mugu, CA 1961. (SECRET)

Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat: Part II," CORG-SP-190, Combat Operations Research Group, HQ US Army Combat

Developments Command, August 1964.

Helmbold, Robert L., "Some Observations on the Use of Lanchester's Theory for Prediction," Operations Research, v. 12(1964), pp. 778-781.

Helmbold, Robert L., "Problems in Testing Combat Models," Presented to the First National Meeting, Military Applications Section, ORSA, Houston, TX, 1965.

Helmbold, Robert L., "A Modification of Lanchester's Equations," Operatons Research, v. 13(1965), pp. 857-859.

Helmbold, Robert L., "Some Observations on Validating Combat Models," Presented to the NATO Conference on Recent Developments In Lanchester Theory, Munich, Germany, 1967.

Helmbold, Robert L., "Probability of Victory in Land Combat as Related to Force Ratio," Rand Paper P-4199, The Rand Corporation, Santa Monica, CA, October 1969.

Helmbold, Robert L., "Have the World Series Been Fixed?", RAND P-4447, The Rand Corporation, 1600 Main St., Santa Monica, CA, August 1970.

Helmbold, Robert L., "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," Rand Report R-772-PR, The Rand Corporation, Santa Monica, CA, June 1971.

Helmbold, Robert L., "Air Battles and Ground Battles -- A Common Pattern?", Rand Paper P-4548, The Rand Corporation, Santa Monica, CA, 1971.

Helmbold, Robert L., "Some Observations on the Choice of Exponent in Lanchester's Law," Presented to the Ad Hoc Working Group on the Theory of Combat, Cosponsored by the Defense Nuclear Agency (DNA) and the Naval Postgraduate School (NPGS), held at NPGS, Monterey, CA, 1979.

Helmbold, Robert L., "An Essay ol Criteria for a Theory of Combat," Presented to a Meeting of the Committe to Develop a Theory of Combat, Naval Postgraduate School, CA, November 1982.

Helmbold, Robert L. (co-author) "EM-1: An Escalation Model Based on Historical Combat Data," (In "Integrated Warfare -- Representing the Decision Processes Within Computer Simulations of Combat," Technical Paper CAA-TP-85-1, US Army Concepts Analysis Agency, Bethesda, MD, January 1985.)

Hempel, Carl G., "Aspects of Scientific Explanation," Free Press, 1965.

HERO, "Historical Trends Related to Weapon Lethality," AD 458 760, 15 Oct 1964.

HERO, "Historical Trends Related to Weapon Lethality: Basic Historical Studies, Volume I," AD 458 757, 15 Oct 1964.

HERO, "Historical Trends Related to Weapon Lethality: Science, Technology, and Weapons Development in History Annex, Volume II," AD 458 758, 15 Oct 1964.

HERO, "Historical Trends Related to Weapon Lethality: Comparative Analysis of Historical Studies Annex, Volume III," AD 458 759, 15 Oct 1964.

HERO, "A Survey of 'Quick Wins' in Modern War," AD-A025 893, Historical Evaluation and Research Organization, Dunn Loring, Virginia, October 1975.

HERO, "Analysis of Factors That Have Influenced the Outcomes of Battles and Wars: A Data Base of Battles and Engagements," Vols. I-VI, CAA-SR-84-6, Sep 84

HERO, "Historical Evaluation of Barrier Effectiveness," AD A050 781. Mar 1974.

HERO, "A Survey of 'Quick Wins' in Modern War," AD A025 893, Oct 1975.

Hibbs, Douglas A., Jr., "Mass Political Violence: A Cross-National Causal Analysis," John Wiley & Sons, New York, 1973.

Hogg, Robert V., "Statistical Robustness: One View of Its Use In Applications Today," The Americal Statistician, v. 33(1979), No. 3(Aug), pp. 108-115.

Holsti, Kal J., "International Politics: A Framework for Analysis," Prentice Hall, Engelwood Cliffs, NJ, 1972.

Holsti, Ole R., "Crisis, Escalation, War," McGill-Queen's University Press, Montreal, Canada, 1972.

Holsti, Ole R., P. T. Hoffman, and John D. Sullivan, "Unity and Disintegration in International Alliances: Comparative Studies," John Wiley & Sons, New York, 1973.

Holsti, Ole R., "Foreign Policy Decision Makers Viewed Psychologically: 'Cognitive Process' Approaches," 1976. (In, "In Search of Global Patterns," James R. Rosenau (Ed), q.v.)

Horvath, William J., and Caxton C. Foster, "Stochastic Models of War Alliances," Journal of Conflict Resolution, v. 7(1963), No. 2(Jun), pp. 110-116.

Horvath, William J., "A Statistical Model for the Duration of Wars and Strikes," Behavioral Science, v.13 (Jan 1968), pp. 18-28.

Huber, Peter J., "Projection Pursuit," Annals of Statistics, Vol 13(1985), No 2(Jun), pp 435-525.

Huber, Reiner K., Lawrence J. Low, and James G. Taylor, "Some Thoughts on Developing a Theory of Combat," NPS 55-79-014, Naval Postgraduate School, Monterey, CA, July 1979.

Hughes, MG B. P., "Firepower: Weapons Effectiveness on the Battlefield, 1630-1850," Charles Scribner's Sons, New York, 1974.

Hulse, BG A. D., "Movement," Combat Operations Research Group, Continental Army Command, Ft. Monroe, VA, 1954.

- I -

- J -

Jacob, Herbert (Ed.), "Using Published Data: Errors and Remedies," Sage Publications, Beverly Hills, CA, 1984.

Jensen, Lloyd, "American Foreign Policy Elites and the Prediction of International Events," Peace Research Society (International) Papers, v. 5(1966), pp. 199-209.

Jessup, John E., and Robert W. Coakley, "A Guide to the Study and Use of Military History," U. S. Army Center for Military History, Washington, DC, 1979.

Johnson, Norman L. and Samuel Kotz, "Distributions in Statistics," John Wiley & Sons, New York, 1970.

Jomini, Baron Henri de, "The Art of War," 1862. (Translated by G. H. Mendell & W. P. Craighill, Lippincott, Philadelphia.)

Jones, Susan D., and J. David Singer, "Beyond Conjecture in International Politics: Abstracts of Data-Based Research," Peacock, Itasca, IL, 1972.

- K -

Kahneman, David, Paul Slovic, and Amos Tversky, "Judgement Under Uncertainty: Heuristics and Biases," Cambridge University Press, Cambridge, 1982.

Kaplan, A., "The Conduct of Inquiry," Chandler, San Francisco, CA, 1964.

Karlowitz, Stan A., "An Experimental Test of a Rationalistic Theory of Deterrence," Journal of Conflict Resolution, v. 17(1973), September, pp. 535-572.

Kelley, C. T., Jr., G. K. Smith, and W. M. Mateskiela, "Quality-Quantity Tradeoffs: A Historical Analysis of Air Combat," Rand Report R-1976-AF, The Rand Corporation, Santa Monica, CA, April 1977.

Kendall, Maurice G., and Alan Stuart, "The Advanced Theory of Statistics," (3 vols), Charles Griffin & Co, Ltd., London, v. 1(1969), v. 2(1967), v. 3(1968).

Kendall, Maurice G., "Multivariate Analysis," Charles Griffin & Co., London, 1975.

Kennedy, Robert F., "Thirteen Days: A Memoir of the Cuban Missile Crisis," Signet Books, New York, 1968.

Kirkpatrick, D. L. I., "Do Lanchester's Equations Adequately Model Real Battles?", RUSI Journal, June 1985.

Klingberg, Frank L., "Predicting the Termination of War," Journal of Conflict Resolution, v. 10(1966), June, pp. 129-171.

Knorr, Klaus, "The War Potential of Nations," Princeton University Press, Princeton, NJ, 1956. (Reprinted 1979 by Greenwood Press, Westport, CT.)

Knorr, Klaus, "Military Power and Potential," D.C. Heath, Lexington, MA, 1970.

Koopman, Bernard O., "An Operational Critique of Detection Laws," Operations Research, v. 27(1979), No. 1(Jan-Feb), pp. 115-133.

Kugler, Jacek, "The Consequences of War: Fluctuations in National Capability Following Major Wars, 1880-1970," University of Michigan, Ann Arbor, MI, 1973.

Kuhn, T., "The Structure of Scientific Revolution," University of Chicago Press, Chicago, IL, 1970.

- L -

Lagerstrom, R. P., and R. C. North, "An Anticipated Gap, Mathematical Model of International Dynamics," Stanford University, Stanford, CA, 1969.

Lanchester, Frederick W., "Aircraft in Warfare: the Dawn of the Third Arm," Constable & Co., Ltd., London, 1916. (The portion dealing with "Lanchester's Square Law" was republished in "The World of Mathematics," Vol. 4, pp. 2138-2157, Edited by James R. Newman, published by Simon & Schuster, New York, 1956).

Latchow, John H., "A Lanchester Model for Air Battles," Report No. GSA/SM/72-8, Air Force Institute of Technology, Wright-Patterson AFB, Dayton, OH, February 1972.

Lawrence, R. J., "The Lognormal Distribution of the Duration of Strikes," J. R. Statist. Soc. A (1984), v. 147, Part 3, pp. 464-483.

Leontieff, Wassily, "Academic Economics," Science, Vol. 217, 9 July 1982, pp. 104-107.

Levy, Jack S., "War in the Modern Great Power System, 1495-1975," University Press of Kentucky, 1983.

Lewin, Kurt, "Field Theory in Social Science," Harper Torchbooks, New York, 1964.

Lider, Julian, "Military Theory: Concept, Structure, Problems," St. Martin's Press, New York, 1982.

Livermore, Thomas L., "Numbers and Losses in the Civil War in America: 1861-1865," Houghton-Mifflin, Boston, MA 1901. (Republished 1957 by Indiana University Press. Reprinted 1977 by Kraus Reprint Co., Millwood, NY.)

Lotka, Alfred J., "Elements of Physical Biology," 1924. Reprinted as "Elements of Mathematical Biology," Dover Publications, New York, 1956.

Love, A. H., "Battle Casualties," U. S. Army Medical Corps, World War 1.

Low, Lawrence, "A Theory of Combat -- A Concept and Plan for Research," SRI International, Menlo Park, CA, 9 April 1980.

Low, Lawrence, "Theater-Level Gaming and Analysis Workshop for Force Planning," held at the Xerox International Center for Training and Management Development, Leesburg, VA on 27-29 Sep 1977. Sponsored by Office of Naval Research Code 431, Arlington, VA 22217. Proceedings published by SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025 under Contract N00014-77-C-0129, May 1981.

Low, William, "Discoveries, Innovations, and Business Cycles," Technological Forecasting and Social Change, Vol 26(1984), No 4(Dec), pp. 355-373.

Lulejian & Associates, Ref 3-16. "A Lanchester Fit to Selected Land Battles," Contract F44620-70-C-0056, February 1971.

- M -

Marshall, S. L. A., "Men Against Fire," William Morrow & Co., New York, 1947.

McGuire, Martin C., "Lanchester Models of Battlefield Attrition and the Organization of Combat Forces," Dept. of Economics, University of Maryland, MD, Feb 1985.

McQuie, Robert, George Cassaday, Robert Chapman, and William Montweiler, "Multivariate Analysis of Combat," PRC-R-1143, Planning Research Corporation, Washington, DC, July 1968.

McQuie, Robert, "Military History and Mathematical Analylsis," Military Review, v. 50(1970), no. 5(May):8-17.

Midlarsky, Manus, "On War: Political Violence in the International System," Free Press, New York, 1975.

Montross, Lynn, "War Through the Ages," Harper & Row, New York, 1960.

Morgenstern, Oscar, "On the Accuracy of Economic Observations," Princeton University Press, Princeton, NJ, Second Edition, 1963.

Morrison, Donald F., "Multivaritate Statistical Methods", McGraw-Hill Book Co., New York, NY, 1967.

Morrison, Samuel E., "History of US Naval Operations in World War II," 15 volumes, Little, Brown & Co., Boston, MA, 1953.

Morse, Philip M., and George E. Kimball, "Methods of Operations Research," John Wiley & Sons, New York, 1951.

Mosteller, Frederick, and John W. Tukey, "Data Analysis and Regression: A Second Course," Addison-Wesley Publishing Co., Reading, MA, 1977.

Mottelay, Paul F., and T. Campbell-Copeland, "The Soldier in Our Civil War," Stanley Bradley Publishing Co., New York/Atlanta, 1893.

Mulholland, R.P., and R.D. Specht, "The Rate of Advance of the Front Line in Some World War II Campaigns," Rand Memorandum RM-1072, The Rand Corporation, Santa Monica, CA, 16 April 1953.

Murphy, COL Joseph B., "Middle East War Game (MEWAG)(U)," CAA-SR-75-1, US Army Concepts Analysis Agency, Bethesda, MD, January 1975. (SECRET-NOFORN)

- N -

Narroll, Raoul, Vern L. Bullogh, and Frada Narroll, "Military Deterrence in History," State University of New York Press, 1974.

Nazareth, MAJ J., and MAJ M. J. W. Wright, "Two Views on the Principles of War," Military Review, US Army Command and General Staff College, February 1961.

Nevins, Allan, James I. Robertson, Jr., and Bell I. Wiley (Eds), "Civil War Books: A Critical Bibliography," Louisiana State University Press, Baton Rouge, LA, 1967.

Nicholson, Michael, "Conflict Analysis," The English Universities Press, Ltd., 1970.

Nie, Norman H., et al, "SPSS Statistical Package for the Social Sciences", Second Edition, McGraw-Hill Book Co., New York, NY, 1975.

North, Robert, Richard Brody, and Ole Holsti, "Some Empirical Data on the Conflict Spiral," Peace Research Society (International) Papers, v. 1(1964), pp. 1-14.

Norwood, COL Fleming W., "analysis of Combat Attrition and Intensity of War (CATIWAR)(U)," CAA-TP-80-4, US Army Concepts Analysis Agency, Bethesda, MD, April 1980. (SECRET-NOFORN)

- O -

O'Leary, Michael K., and Coplin, William D., "Quantitative Techniques in Foreign Policy Analysis and Forecasting," Praeger Publishers, New York, 1975.

Oman, Charles, "The Art of War in the Sixteenth Century," E. P. Dutton & Co., New York, 1937.

Organski, A. F. K., and Jacek Kugler, "The War Ledger," University of Chicago Press, Chicago, IL, 1980.

Osipov, M., "The Influence of the Numerical Strength of Engaged Sides on Their Losses," (in Russian). A five-part article published in Voenniy Sbornik, 1915, as follows. Part 1, No. 6(Jun), pp. 31-40; Part 2, No. 7(Jul), pp. 25-36; Part 3, No. 8(Aug), pp. 31-40; Part 4, No. 9(Sep), pp. 25-37; Part 5, No. 10(Oct), pp. 93-96.

Ostermann, L. E., "The Variation of Loss Ratios in Combat," Strategic and Tactical Studies, Lulejian and Associates, Redondo Beach, CA Dec 1972.

Overholt, J. L., "Fitting Korean War Data by Statistical Methods," Professional Paper No. 32 (AD 705 349), The Center for Naval Analyses, Arlington, VA, May 1970.

Overholt, J. L., et al, "Study of Land/Air Tradeoffs, Volume IX: Statistical Analysis of Korean War Data -- Spring 1951," Study No. 64 (AD 869 481), The Center for Naval Analyses, Arlington, VA, Apr 1970.

- P -

Pagano, Judy K., and W. E. Sykes, "For Want of a Theory," Presented to the Second Working Meeting of the Ad Hoc Committee to Develop a Theory of Combat

Held at the Defense Nuclear Agency, Alexandria, VA 28-29 January 1980.

Parsons, Col Norman W., "Military Unit Rates of Advance," Combat Operations Research Group, Fort Monroe, Virginia, 1954.

Patton, Gen George S., Jr., "War As I Knew It," Bantam Publications, New York, NY, 1947.

Peterson, Richard H., "On the 'Logarithmic Law' of Attrition and Its Application to Tank Combat," Operations Research, v. 15(1967), no. 3(May-Jun), pp. 557-558.

Philips, T.R. (Ed.), "Roots of Strategy," Military Service Publishing Co., Harrisburg, PA, 1940.

Phisterer, Frederick, "Statistical Record of the Armies of the United States: Supplementary Volume to Campaigns of the Civil War," Charles Scribner's Sons, New York, 1883.

Platt, John R., "Strong Inference," Science, v. 146(1964), No. 3642(16 October), pp. 347-353.

Poincare, Henri, "Science and Hypothesis," Dover Publications, New York, 1952.

Popper, K. R., "The Logic of Scientific Discovery," Science Publications, New York, 1961.

Posony, Stefan T., "Waking Up the Giant," Arlington House, 1974.

Pratt, John W., and Robert Schlaifer, "On the Nature and Discovery of Structure," Journal of the American Statistical Association, v. 79(1984), No. 385(Mar), pp. 29-33.

- Q -

- R -

Raiffa, Howard, "Decision Analysis," Addision Wesley, Reading, Massachusetts, July 1970.

Rao, C. Radhakrishna, "Linear Statistical Inference and Its Applications," John Wiley & Sons, New York, 1965. (Second Edition 1973).

Rapoport, Anatol, "Fights, Games, and Debates," University of Michigan Press, Ann Arbor, MI, 1960.

Rashevsky, Nicholas, "Mathematical Theory of Human Relations: An Approach to a Mathematical Biology of Social Phenomena," (reprint) The Principia Press, Bloomington, IN, 1974.

Reister, Frank A., "Medical Statistics in World War II," Office of The Surgeon General, Department of the Army, Washington, DC 1975.

Richardson, Lewis F., "Arms and Insecurity," Boxwood and Quadrangle Press, Pittsburgh and Chicago, 1960.

Richardson, Lewis F., "Statistics of Deadly Quarrels," Boxwood and Quadrangle Press, Pittsburgh and Chicago, 1960. (3-rd printing by the Boxwood Press, Pacific Grove, CA, 1975.)

Richardson, R. P., et al, "An Analysis of Recent Conflicts," CNA Research Contribution No. 144, Center for Naval Analyses, Alexandria, VA, January 1966.

Robinson, Julian Perry, "Neutron Bomb and Conventional Weapons of Mass Destruction," Bulletin of the Atomic Scientists, March 1978, pp. 42-45.

Rosenau, James N. (Ed.) "In Search of Global Patterns," The Free Press, New York, 1976.

Rosen, Steven, "Tolerance of Human Life Costs for Foreign Policy Goals," Peace Research Society (International) Papers, v. 17(1971), pp. 61-73.

Rosenbaum, Paul R., "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment," Journal of the American Statistical Association, v. 79(1984), No. 385 (Mar), pp. 41-48.

Rowny, Don Karl (Ed.), "Soviet Quantitative History," Sage Publications, Beverly Hills, CA, 1984.

Rubin, Ernest, "Tournament and Match Play in Chess," The American Statistician, June 1969, pp. 43-46.

Rummel, R. J., "Dimensions of Nations," Sage Publications, Beverly Hills, CA, 1972.

Russett, Bruce M., et al, "World Handbook of Political and Social Indicators," Yale University Press, New Haven, CT, 1964. (Second Edition -- see listing under the name of Taylor, Charles L.)

Russett, Bruce M., J. D. Singer, and M. Small, "National Political Units in the 20-th Century: A Standardized List," American Political Science Review, v. 62 (1968), No. 3, pp. 932-951.

Russett, Bruce M., "Peace, War and Numbers," Sage Publications, Beverly Hills, CA, 1972.

- S -

Samz, Robert W., "Some Comments on Engel's 'A Verification of Lanchester's Law'," Operations Research, v. 20(1972), pp. 49-52.

Saperstein, Alvin M., "Chaos -- A Model for the Outbreak of War," Nature, v. 309, pp. 303-305, 24 May 1984.

Savage, Leonard J., "Foundations of Statistics," Dover Publications, New York, 1972. (Originally published by John Wiley & Sons, New York, 1954.)

Savkin, Vasilii E., "Rates of Advance," (in Russian), Military Publishing House, Ministry of Defense, Moscow, 1965.

Saxe, Marshall Maurice de "Reveries on the Art of War," 1757.

Schelling, Thomas C., "The Strategy of Conflict," Harvard University, Cambridge, MA, 1960.

Schmieman, William A., "The Use of Lanchester-Type Equations in the Analysis of Past Military Engagements," Doctoral Thesis, Georgia Institute of Technology, August 1967.

Schrodt, Philip A., "Microcomputers in the Study of Politics: Predicting Wars With the Richarson Arms-Race Model," BYTE Magazine, July 1982, pp. 108-134.

Selye, Hans (Ed), "Selye's Guide to Stress Research," (3 vols), Van Nostrand Reinhold, Florence, KY, 1980-1983

Shubik, Martin, and G. D. Brewer, "Models, Simulations and Games -- A Survey," Rand Report R-1060-ARPA/RC, The Rand Corporation, Santa Monica, CA, May 1972.

Simon, Herbert A., "Prediction and Hindsight as Confirmatory Evidence," Philosophy of Science, v. 22(1955), pp. 227-230.

Singer, J. David, "Inter-Nation Influences: A Formal Model," American Political Science Review, v. 57(1963), No. 2(June), pp. 420-430.

Singer, J. David (Ed.), "Quantitative International Politics: Insights and Evidence," Free Press, New York, 1968.

Singer, J. David, and Melvin Small, "The Wages of War 1816-1965: A Statistical Handbook," John Wiley & Sons, Inc., New York, 1972.

Singer, J. David and Michael D. Wallace (Eds.), "To Augur Well: Early Warning Kdicators in World Politics," Sage Publications, Inc., Beverly Hills, CA, 1979.

Singer, J. David (Ed.), "The Correlates of War I: Research Origins and Rationale," The Free Press, Macmillan Publishing Co., New York, 1980.

Singer, J. David (Ed.), "The Correlates of War: II -- Testing Some Realpolitik Models," The Free Press, Macmillan Publishing Co., New York, 1980.

Small, Melvin, and J. David Singer, "Resort to Arms: International and Civil Wars, 1816-1980," Sage Publications, Beverly Hills, CA, 1982.

Smith, E. G, and G. N. Donovan, "Trends in Warfare," Memorandum No. F-6, British Army Operational Research Group, April 1955. (RESTRICTED)

Smoker, Paul, "An International Process Simulation: Development, Usage and Partial Validation. Part I: International Processes Simulation, A Man-Computer Model. Part II: Analysis of Conflict Behaviors in an International Processes Simulation and International System, 1955-1960," University Microfilm, Ann Arbor, MI, 1973.

Snedecor, George W., and William G. Cochran, "Statistical Methods", Fifth Edition, The Iowa State University Press, Ames, IA, 1961.

Sorokin, Pitirim A., "Social and Cultural Dynamics," (4 vols.), 1937-1941. (Reissued 1962.)

Spaulding, Oliver L., Hoffman Nickerson, and John W. Wright, "Warfare," The Infantry Journal, Inc., Washington, DC, 1937.

Strauch, Ralph E., "A Critical Assessment of Quantitative Methodology As a Policy Analysis Tool," P-5282, The RAND Corporation, Santa Monica, CA, Aug 1974.

Steele, Matthew W., "American Campaigns," Office Chief of Staff, War Department, 1909. (Republished by Military Service Publishing Co., Harrisburg, PA, 6-th printing 1949.)

Stockfisch, J. A., "Models, Data and War: A Critique of the Study of Conventional Forces," Rand Report R-1526-PR, The Rand Corporation, Santa Monica, CA, March 1975.

Sun Tzu, "The Art of War," 500 BC. Translated by Samuel B. Griffith, The Clarendon Press, Oxford, 11-th printing 1982.

Sutherland, R. E., "Some Basic Problems in War Gaming of Ground Combat," RAC-TP-153 (AD 470 102), Research Analysis Corporation, McLean, VA, Mar 1965.

Sykes, W. E., and Judy K. Pagano, "Models, Data, and War: A Critique of the Foundations of Defense Analyses," PAD-80-12, U. S. General Accounting Office, 12 March 1980.

Taylor, Charles L., and Michael C. Hudson, "World Handbook of Political and Social Indicators," (Second Edition), Yale University Press, New Haven, CT, 1972. (First Edition -- see listing under the name of Russett, Bruce M.)

Taylor, James G., "Lanchester Models of Warfare," Vols. I & II, Military Applications Section, Operations Research Society of America, Arlington, VA, 1983.

Thompson, William R., and L. Gary Zuk, "War, Inflation, and the Kondratieff Long Wave," Journal of Conflict Resolution, Vol 26(1982), No 4(Dec), pp. 621-644.

Thorp, K., et al, "Comparison of the CARMONETTE Model with the TETAM Field Experiment," CAA-TP-75-8 (AD B008 161), US Army Concepts Analysis Agency, Bethesda, MD, Dec 1975.

Thurstone, L., "Vectors of the Mind," University of Chicago Press, Chicago, IL, 1935.

Timenes, Nikolai, Jr., "Defense Against Kamikaze Attacks in World War II and Its Relevance to Anti-ship Missile Defense," Operations Evaluation Group Study 741, Center for Naval Analyses, November 1970. Vol. 1: "An Analytical History of Kamikaze Attacks Against Ships of the United States Navy During World War II."

Towle, Philip (Ed.), "Estimating Foreign Military Power," Holmes & Meier Publishers, Inc., 1982.

Turney-High, Harry Holbert, "Primitive War: Its Practice and Concepts," University of Southern Carolina Press, Columbia, SC, 1949. (Second Edition 1971.)

Tryfos, P., S. Casey, S. Cook, G. Leger, and B. Pylpiak, "The Profitability of Wagering on NFL Games," Management Science, v. 30(1984), No. 1(Jan), pp. 123-132.

Tse-Tung, Mao, "On Guerilla Warfare," 1937. Translated by Samuel B. Griffith, Frederick A. Praeger, Inc., 1961.

- U -

- V -

Vagts, Alfred, "Battle and Other Combatant Casualties in the Second World War," Part 1, pp. 256-294; Part 2, pp. 411-438; J. Politics, v. 7(1945).

Vagts, Alfred, "Landing Operations," Military Service Publishing Co., Harrisburg, PA, 1946.

Verhulst, Michael, "The Concept of a 'Mission'," Naval Research Logistics Quarterly, v.3(1956), pp. 45-57.

Vickers, Brian (Ed), "Occult and Scientific Mentalities in the Renaissance," Cambridge Univeristy Press, New York, June 1982.

Voevodsky, John, "A Mathematical Representation of Time-Variant Quantitative Behavior," Journal of Psychology, v. 68(1968), pp. 129-140.

Voevodsky, John, "Quantitative Behavior of Warring Nations," Journal of Psychology, v. 72(1969), pp. 269-292.

Voevodsky, John, and S. L. Waleszczak, "Quantitative Analysis of the Vietnam War Through 30 June 1969," Submitted to the Military Operations Research Symposium for General Sessions Consideration, 27 February 1970.

Voevodsky, John, "Crisis Waves: The Growth and Decline of War-Related Behavioral Events," Journal of Psychology, v. 80(1972), pp. 289-308.

- W -

Wainstein, L., "An Examination of the Parsons and Hulse Papers on Rates of Advance," Institute for Defense Analyses, Dec 1973.

Wainstein, L., "Rates of Advance in Infantry Division Attacks in the Normandy-Northern France and Siegfried Line Campaigns," P-990 (AD 779 882), Institute for Defense Analyses, Arlington, VA, Dec 1973.

Wainstein, L., "Some Allied and German Casualty Rates in the European Theater of Operations," P-989 (AD 779 850), Institute for Defense Analyses, Arlington, VA, Dec 1973.

Wallace, Michael, "War and Rank Among Nations," Heath, Lexington, MA, 1973.

Ward, Michael D., "Simulating the Arms Race," BYTE, October 1985, pp. 213-222.

Weiss, Herbert K., "Requirements for a Theory of Combat," Memorandum Report No. 667, Ballistic Research Laboratories, Aberdeen Proving Ground, MD, April 1953.

Weiss, Herbert K., "Some Mathematical Models of Tank Combat," BRL-MR-724, US Army Ballistic Research Laboratories, Aberdeen, MD, September 1953.

Weiss, Herbert K., "Lanchester-Type Models of Warfare," in "Proceedings of the First International Conference on Operations Research," Operations Research Society of America, 1957.

Weiss, Herbert K., "Foreseeable Changes in Operations Research Tasks, Techniques, and Organizations," Published as a Chapter in "Proceedings of the 20-th Anniversary Conference of the Operations Evaluation Group," 14-16 May 1962.

Weiss, Herbert K., "Influence of the Guidance Designer on Warfare," Published as a chapter in "Air, Space, and Instruments," Sidney Lees (Ed.), Draper Anniversary Volume, McGraw-Hill, New York, 1963

Weiss, Herbert K., "Stochastic Models for the Duration and Magnitude of a 'Deadly Quarrel'," Operations Research, v. 11(1963), pp. 101-121.

Weiss, Herbert K., "Combat Models and Historical Data: The U.S. Civil War," Operations Research, v. 14(1966), pp. 759-790.

Westergaard, Harald (Ed.), "Losses of Life Caused by War," Clarendon Press, Oxford, 1923.

Westfall, Richard S., "Newton and the Fudge Factor," Science, v. 179(1973), no. 4075 (23 Feb): 751-758.

Wilkenfeld, Jonathon (Ed.), "Conflict Behavior and Linkage Politics," David McKay Company, Inc., 1973.

Wilks, Samuel S., "Mathematical Statistics," John Wiley & Sons, New York, 1962.

Willard, Daniel, "Lanchester as Force in History: An Analysis of Land Battles of the Year 1618-1905," RAC-TP-74, Research Analysis Corporation, November 1962.

Williams, W. H., "How Bad Can 'Good' Data Really Be?", The American Statistician, v. 32(1978), No. 2(May), pp. 61-65.

Willis, Roger, "Potential Uses for Historical Combat Data and Information," OR/WP/16, The Royal Military College of Science, Shrivenham, UK, Dec 1982.

Willoughby, Charles, "Maneuver in War," Military Service Publishing Co, Harrisburg, PA, 1939. (Reprinted November 1983 by US Army War College, Carlisle, PA.)

Wilson, CAPT A. H., and CDR W. R. Finn, "Improvise or Plan?", Naval Research Logistics Quarterly, v.4(1957), 263-267.

Woods, F. A., and A. Baltzly, "Is War Diminishing?", Houghton Mifflin, Boston, MA, 1915.

Woytinsky, W. L., "Die Welt in Zahlen," Berlin, 1928.

Wright, Quincy, "A Study of International Relations," Appleton-Century-Crofts, New York, 1955.

Wright, Quincy, "A Study of War," (2 Vols) University of Chicago Press, 1942. 1-Volume 2-nd edition, 1965. Second impression 1967.

Wu, Teh-Wei, "Econometrics: An Introductory Analysis", University Park Press, Baltimore, MD, 1973.

- X -

- Y -

- Z -

Zupan, Jure, "Clustering of Large Data Sets," Research Studies Press, A Division of John Wiley & Sons, Ltd., Chichester, England, 1982.

APPENDIX E

DEFINITIONS AND ABBREVIATIONS USED IN THE NONCOMPUTERIZED VERSION OF THE HERO DATA BASE

E-1. ORGANIZATION

a. The battles and engagements of the HERO data base are divided chronologically into five approximately equal groups, defined by the following time periods:

(1) 17th and 18th Centuries (1600-1800; Volume II)

(2) 19th Century (1805-1900; Volume III)

(3) Early 20th Century (1904-1940; Volume IV)

(4) Mid-20th Century to 1945 (1939-1945; Volume V)

(5) 20th Century since 1939 (1939-1973; Volume VI)

b. Within each time period, major wars are listed, and within each war significant details of a number of selected battles and engagements are presented. In the cases of wars from which only a few engagements appear in the HERO data bases, all these engagements are often grouped together, primarily for organizational simplicity.

c. For each major war, or group of wars, the HERO data base provides in tabular form a summary of the important numerical data and qualitative information concerning each battle, plus a historical assessment of the factors that were important to the outcome of the battle. Following each such table or group of tables are narrative summaries of the battles listed in the table(s). These narrative summaries include a brief assessment of the significance of the battle, and also identify the sources consulted with respect to the presentation for that battle.

d. Discussed below are the significant definitions for each of the seven major tables of the HERO data base, as well as the abbreviations and symbols used for the original noncomputerized presentation of the data.

E-2. DEFINITIONS. All terms defined below were introduced and used by HERO to characterize the nature and outcomes of the various engagements in their data base.

a. Table 1 - Identification

(1) Engagement. In the HERO data base this term is used in a broad sense and comprehends significant combat encounters between hostile forces at various levels of aggregation from small unit up to and including corps, Army, and Army group. The descriptor used in each case provides the engagement name and (in Table 1 only) the geopolitical area in which the engagement took place.

(2) Dates. The date on which a particular engagement began.

(3) Campaign. The recognized or appropriate designation for a connected series of military operations forming a distinct stage in a war.

(4) War. A contest by military force, involving extreme violence, waged between two or more nations, states, or other politically organized bodies.

(5) Attacker. That military force which, at the beginning or in the first phase of an engagement, initiates and sustains significant offensive action against its opponent.

(6) Defender. That force which, at the outset or in the first phase of an engagement, chooses to maintain or is forced to adopt a defensive posture.

(7) Attacker CO. The officer or general officer who exercises command over the attacking force.

(8) Defender CO. The officer or general officer who exercises command over the defending force.

(9) Duration. The extent of time, expressed in number of days, during which an engagement takes place. For purposes of this report, a portion of a day is considered a full day. The sole (and logical) exception to this rule occurs in cases of overnight engagements in which significant combat began in the late afternoon or evening of one day and was concluded before noon of the following day. In such cases the engagements are considered one-day engagements, since the duration was less than 24 hours.

(10) Width of Front. The space from side to side or flank to flank occupied or covered by a force just before the onset of the engagement. This distance is measured in kilometers, the measurement generally following the front and ignoring minor salients or reentrants. Where there is a significant difference between the fronts occupied by the opposing forces in an engagement, the width of the attacker's front is entered as the descriptor.

b. Table 2 - Operational and Environmental Variables

(1) Defender Posture. The level of resistance to, or protection from, any and all forms of enemy attack. Five basic levels are identified for purposes of this study:

(a) **Hasty defense:** A defense normally organized while in contact with the enemy or when contact is imminent and time available for organization is limited. It is characterized by improvement of the natural defensive strength of the terrain by utilization of foxholes, emplacements, and obstacles; if occupied for a protracted period the hasty defense position can be improved to the status of prepared or fortified defense.

(b) **Prepared defense:** A defense system prepared by a defender who has had time to organize the defensive position, but which (due to lack of time or resources) has less than the strength of a fortified position.

(c) **Fortified defense:** A comprehensive, coordinated defense system prepared by a defender with sufficient time to complete planned entrenchments, field fortifications, and obstacles in such a manner as to permit the most effective possible employment of defensive firepower.

(d) **Delay (delaying action):** A retrograde movement in which, in successive positions, the defender inflicts maximum delay and damage on an advancing enemy to gain time, without becoming decisively engaged in combat or being outflanked.

(e) **Withdrawal from action:** A retrograde maneuver whereby a force disengages from combat, or contact with an enemy force, in accordance with the will of its commander.

Frequently, it should be noted, descriptors entered in this category reflect a defensive posture best defined as a combination or average of two of the five basic categories. For example, a defender may adopt two postures during the course of an engagement, or the level of defensive preparation may not be uniform across a lengthy front or throughout the depth of a defended zone.

(2) **Terrain.** The nature of the ground on which the engagement was fought, described by its most prominent characteristics.

(3) **Weather.** The meteorological conditions prevailing at the time of the engagement, described generally.

(4) **Season.** The season during which the engagement took place: spring, summer, fall, or winter. This descriptor is valuable principally for providing a rough measure of the hours of daylight available for the employment of weapons.

(5) **Surprise.** For each engagement considered, a determination was made as to whether or not surprise had been achieved by one side or the other, and if it had been, by whom and to what degree.

(a) Surprise, as used in the HERO data base, is defined as a condition which comes into existence when one military force (or its commander) is able to confront the opponent with circumstances that the opponent did not anticipate or adequately provide for. Surprise may be achieved with respect to time, place, or performance.

(b) For this data base, three degrees of surprise were posited: complete, substantial, and minor. Assessments of the degree of surprise achieved were subjective military historical judgments based on the historical record.

(6) **Air Superiority.** This factor is applied only to engagements of World War I (where applicable) and later. It identifies the side whose air force has established a degree of capability over the opposing air force which permits it to conduct air operations at the time and place of the engagement without prohibitive interference from the opposing air force.

c. **Table 3 - Strengths and Combat Outcomes.** This table presents, for attacker and defender, quantitative descriptors of personnel strengths, battle casualties, and, for major items of materiel, strengths and losses. Finally, the table shows the distance advanced, in kilometers, on a per day basis.

(1) **Strength.** This category provides, where appropriate or known, data on the personnel and major materiel strengths of the opposing forces.

(a) **Total (personnel).** The sum, at the start of an engagement, of all personnel subject to enemy fire, including generally combat and combat support troops but also service troops if subject to enemy fire. For lengthy engagements in which both sides were significantly reinforced after the beginning of the engagement, an average of the daily start strength(s) was entered.

(b) **Cavalry.** The number of mounted troops, including dragoons and mounted infantry, at the start of the engagement. This category was employed for engagements prior to World War I.

(c) **Artillery Pieces.** Complete projectile-firing weapons, including cannon, artillery mortars, and multiple rocket launchers.

(d) **Armor.** Armored track-laying vehicles mounting a cannon-type weapon. In this report the armor total includes tanks, armored, self-propelled antitank guns, and armored assault guns, such as the World War II German sturmgeschutz. Where the available data permit, the armor total is further broken down according to whether the armored vehicles employed were light or MBT (i.e., main battle tank). This breakdown was made according to the standards or nomenclature employed by the user force. In the absence of such guidance, the following criteria were employed to differentiate between the two categories:

1. **Light.** Includes armored fighting vehicles (AFVs) up to 25 tons in weight, usually fast and mobile, with primary missions of security and reconnaissance. Does not include armored cars, halftracks, infantry carriers, and armored infantry fighting vehicles.

2. **MBT.** Armored fighting vehicles over 26 tons in weight; including, generally, the principal AFV of armored divisions with the primary mission of engaging and defeating the enemy's armor, all self propelled antitank guns, and all armored assault guns.

(e) **Air Sorties.** The number of single-aircraft missions flown by aircraft against enemy targets in the engagement area. The number includes sorties by fighter, fighter bomber, and bomber aircraft.

(2) **Battle Casualties.** The number of personnel killed, wounded, or missing (including prisoners) during the engagement. Does not include personnel losses resulting from illness, disease, or nonbattle injuries. Battle casualties are entered as the arithmetical total over the course of the engagement (not including prisoners taken in pursuit following the termination of significant combat) and as a figure representing percent per day casualties.

(3) **Artillery Pieces Lost.** Artillery pieces destroyed, damaged (i.e., out of action for at least one day), or captured as a result of enemy action. Such losses are entered as an arithmetical total and as a figure representing percent per day losses.

(4) **Armor Losses.** Tanks and other AFVs (according to the definition above) destroyed, damaged, or captured as a result of enemy action. Such losses are entered as an arithmetical total and as a figure representing percent per day losses.

(5) **Aircraft Losses.** Combat aircraft lost as a result of enemy action. Such losses are represented as an arithmetical total and as a figure representing aircraft losses calculated on a percent sorties per day basis.

(6) In all the above cases involving enumerations or figures, instances in which a number is not known or is not ascertainable from the historical record are indicated by a "?". In such cases it was not possible to calculate percent per day or percent per sortie rates for casualties and materiel losses (or no loss occurred); in these cases the use of a dash ("-") indicates the absence of a calculable figure. The same system applies to calculations of advance rates, although in this case the use of a dash indicates that the defender had no measurable advance.

d. Table 4 - Intangible Factors (Indicators). For each of these factors, judgments based on the military historical record are made. These judgments assess, with respect to the attacker and defender in each engagement, whether the factor was:

- Comparable for both sides
- No factor

- Advantage

- Disadvantage

(1) **Combat Effectiveness.** A complex factor, subsuming--among other elements--leadership, training and experience, morale, and logistics.

(2) **Leadership.** The art of influencing others to cooperate to achieve a common goal, including, for military leaders at all command strata, tactical competence, and initiative.

(3) **Training and Experience.** Training: the relative adequacy of instruction and preparation to meet the exigencies of campaign and combat. Experience: the relative amount of time spent under field and combat conditions, thus gaining knowledge, skills, and techniques otherwise unavailable.

(4) **Morale.** Prevailing mood and spirit conducive to willing and dependable performance, steadiness, self-control, and courageous, determined conduct despite danger and privations.

(5) **Logistics.** Supply capability.

(6) **Momentum.** An advantage comprised of both space and time factors and having to do with impetus.

(7) **Intelligence.** Information about the organization, dispositions, intentions, and activities of the forces of the opponent.

(8) **Technology.** The application of scientific knowledge, methods, or research to the art of warfare.

(9) **Initiative.** An advantage gained by acting first, and thus forcing the opponent to respond to one's own plans and actions, instead of being able to follow his own plans.

e. Table 5 - Outcome. This table provides assessments of combat outcomes in three categories: victor, distance advanced, and mission accomplishment. The definitions of these categories are:

(1) **Victor.** The victor, if not apparent from the decisive resolution of the combat in favor of one side or the other, is determined by an assessment of the extent to which each side was successful in accomplishing its mission. In many engagements, neither side can be designated the victor. Success is designated by the entry of an "x" in the line for attacker or defender. In drawn battles or battles in which both sides attained success, an "x" is entered for both attacker and defender.

(2) **Distance Advanced.** The distance, in kilometers, from the line of departure to the farthest point reached by significant maneuver elements of the attacking force, measured along the axis of advance. The distance advanced, if negligible, is indicated by an "N"; if unknown or not ascertainable from the record, it is indicated by a "?".

(3) **Mission Accomplishment.** The numerical score on a scale of 0-10 indicates the extent to which each force was successful in accomplishing its mission. Higher scores are given to greater success. The score was determined by the use of HERO's Mission Accomplishment Worksheet, a score sheet which allows the assignment of quantitative values of from 0-2 in each of five categories determined to indicate the relative success or failure of a force in accomplishing its mission during an engagement. The scores awarded in each category are totalled to give the total mission accomplishment score. Scores assigned are the result of the application of experienced subjective military historical judgment. Occasionally, penalties or bonus points may be deducted or awarded for extraordinarily poor or good performances in one or more of the five categories. Definitions of the five elements of mission accomplishment follow:

(a) **Conceptual Accomplishment.** The relative success or failure of the force in executing the operational plan of the commander.

(b) **Geographical Accomplishment.** The relative success or failure of the force in taking or holding positions or position areas in conformity with the operational plan of the commander.

(c) **Prevent Hostile Mission.** The relative success or failure of a force in denying to the enemy the fulfillment of his objectives.

(d) **Command and Staff Performance.** An evaluation of the efficiency and efficacy of the decisions made and actions taken by the officers in command and staff positions in connection with the onset, course, and outcome of an engagement.

(e) **Troop Performance.** An evaluation of the overall combat efficiency and effectiveness of the troops engaged in the course of an engagement.

f. **Table 6 - Factors Affecting Outcome.** Here are listed those factors, tangible and intangible, that seem to have had particular effect upon battle outcomes; the extent to which these are relevant in each battle is indicated. The factors are:

(1) **Force Quality.** The relative combat capability of the forces engaged, including the quality of lower-level and intermediate leadership, but not that of top leadership, which is considered to be a discrete factor.

(2) **Reserves.** The extent to which reserves were available and were committed in a timely manner.

(3) **Mobility Superiority.** The relative quality or numbers of mounted forces, whether horse, horse-drawn, or automotive, expressed in terms of tanks and armored and unarmored vehicles.

(4) **Air Superiority.** The effect one force's command of the air space above the battlefield, if present, had on the outcome of the engagement.

(5) **Terrain, Roads.** The extent to which terrain considerations affected one side to a significantly greater extent than the other.

(6) **Leadership.** The relative capability of top leadership.

(7) **Planning.** The relative effectiveness of prebattle plans and preparations.

(8) **Surprise.** How surprise, if present, aided one side or the other.

(9) **Maneuver.** The effect of a commander's decision, and action implementing the decision, to position his forces for optimum effectiveness in accomplishing his mission, to include the massing of forces on a narrow front.

(10) **Logistics.** The extent to which logistics influenced a battle, remembering that the effects of logistics usually affect a campaign, rather than a single battle.

(11) **Fortifications.** The influence of a defender's fortifications.

(12) **Depth.** The impact of either the attacker or defender being arrayed in depth.

(13) **Weather and Force Preponderance.** These factors, although listed in the data tables, were not explicitly defined by HERO in **Ref 1-1.** Presumably, they represent the effects of these factors on the outcome of the battle.

g. **Table 7 - Combat Forms and Resolution.** This table permits representation, through symbols and abbreviations, of the general nature of the combat in a battle, in terms of force dispositions and maneuver, plus representation of the outcome and immediate after-effects of the battle or engagement. This is shown in terms of the following:

(1) **Main attack and scheme of defense.** Abbreviations show various forms of deployment and maneuver of both sides.

(2) **Secondary attack.** This is shown in the same fashion.

(3) **Success.** Indicates which side was successful.

(4) **Resolution.** Shows what happened to both sides as a result of the battle.

E-3. ABBREVIATIONS AND SYMBOLS. A system of abbreviations and symbols is used for table entries. These are shown below.

a. Table 1 - Identification. The symbols used in this table are as follows:

Am	American
Amph	Amphibious
Armd	Armored
Aus	Austrian
Bav	Bavarian
Bde	Brigade
Bn	Battalion
Boer	Boer
Boh	Bohemian
Br	British
Br Exped Force	Expeditionary Force
Brig	Brigadier
Brig Gen	Brigadier General
Bul	Bulgarian
Cav	Cavalry
Col	Colonel
Cov	Scots Covenanter
CCA	Combat Command A
CCB	Combat Command B
CCR	Combat Command Reserve
CG	Commanding General
Co	Company
Cos	Companies
Cr Pr	Crown Prince
CS	Confederate States (of America)
Cumb'd	Cumberland
Dan	Danish
Det	Detachment
DK	Duke
Du	Dutch
Eg	Egyptian
elms	Elements
Eng	English
Eth	Ethiopian
Fld	Field
FM	Field Marshal
Ft Rgt	Foot Regiment
Fr	French
Ger	German
Gds	Guards
Gr	Grenadier
Han	Hanoverian
Imp	Imperialist
Ind Inf Bn	Independent Infantry Battalion (Japanese)
Is	Israeli

It	Italian
Jap	Japanese
Jgr	Jaeger
Jor	Jordanian
KG	Kampfgruppe (German combat term)
Mam	Mameluke
Mar	Marine
Mech	Mechanized
Mes	Mesopotamian
Mex	Mexican
MG	Major General
Para	Paratroop
Parl	Parliament
PG	Panzer Grenadier
Pied	Piedmontese (Piedmont-Savoy or Piedmont-Sardinia)
PLA	Palestine Liberation Army
Pol	Polish
Port	Portuguese
Pr	Prussian
Prot	Protestant
Reb	Rebel
Res	Reserve
Rgt	Regiment
Rom	Romanian
Roy	Royalist
Russ	Russian
Sax	Saxon
Serb	Serbian
Sp	Spanish
Sp Rep	Spanish Republican
Spec Estab Rgt	Special Established Regiment (Japanese)
Sov	Soviet
Sw	Swedish
Syr	Syrian
TF	Task Force
Tk	Turk
U/I	Unidentified (unit)
US	United States
VG	Volks Grenadier
Vol	Volunteers
(+)	Reinforced
(-)	Elements, part, or a portion of a unit.

b. Table 2 - Operational and Environmental Variables

Defender Posture:

HD	Hasty defense
PD	Prepared defense
FD	Fortified defense
WDL	Withdrawal
Del	Delay

Terrain:

RD	Rolling, desert
RgB	Rugged, bare
RgM	Rugged, mixed
RgW	Rugged, heavily wooded
RB	Rolling, bare
RM	Rolling, mixed
RW	Rolling, heavily wooded
FB	Flat, bare
FM	Flat, mixed
FW	Flat, heavily wooded
FD	Flat, desert
R Dunes	Rolling dunes
U	Urban or built-up area
M	Marsh or swamp

Weather:

DSH	Dry, sunshine, hot
DST	Dry, sunshine, temperate
DSC	Dry, sunshine, cold
DOH	Dry, overcast, hot
DOT	Dry, overcast, temperate
DOC	Dry, overcast, cold
WLH	Wet, light, hot
WLT	Wet, light, temperate
WLC	Wet, light, cold
WHH	Wet, heavy, hot
WHT	Wet, heavy, temperate
WHC	Wet, heavy, cold

Seasons:

Months	**Northern hemisphere**	**Southern hemisphere**
March, April, May	Spring	Fall
June, July, August	Summer	Winter
September, October, November	Fall	Spring
December, January, February	Winter	Summer

Season Codes:

SpT	Spring, temperate
ST	Summer, temperate
FT	Fall, temperate
WT	Winter, temperate
SpTr	Spring, tropical
STr	Summer, tropical
FTr	Fall, tropical
WTr	Winter, tropical
SpD	Spring, desert
SD	Summer, desert
FD	Fall, desert
WD	Winter, desert

Surprise:

Y	Surprise achieved.
N	Surprise did not influence outcome of battle.
X	Symbol showing which side achieved surprise.

c. Table 4 - Intangible Factors

C	Comparable for both sides
N	Not a factor
X	Advantage
O	Disadvantage

d. Table 5 - Outcome

X	Designates successful side
N	Negligible advance

e. Table 6 - Factors affecting Outcome

Same as for Table 4, with the following additions:

X	Advantage decisively affecting outcome
O	Disadvantage decisively affecting outcome

f. Table 7 - Combat Forms and Resolution

Main attack plan and scheme of defense:

F	Frontal attack
E	Single envelopment
EE	Double envelopment
FE	Feint, demonstration, or holding attack
D	Defensive plan
D/O	Defensive-offensive plan
(LF)	Left flank
(RF)	Right flank
(LR)	Left flank and/or rear
(RR)	Right flank and/or rear
P	Penetration
RivC	River crossing
--	No secondary attack

Success: Indicated by an "X"

Resolution:

S	Stalemate
R	Repulse
P	Penetration
B	Breakthrough
WD	Withdrew
WDL	Withdrew with serious loss
A	Annihilated
Ps	Pursued

APPENDIX F

CODING SCHEME FOR THE COMPUTERIZED VERSION OF THE HERO DATA BASE

F-1. NOTES

a. The abbreviations listed below under the heading "Abbrv" are those used in the noncomputerized version of the HERO data base described in Appendix E. The codes listed below are those used in the computerized version of the data base.

b. NN is the total number of battles in the data base. Its FORTRAN format is INTEGER. (For the HERO data base, NN is equal to 601.)

F-2. BATTLE SEQUENCE NUMBER

ISEQNO Three-digit sequence number assigned (by CAA) to the battles in a data base. This sequence number runs from 1 to NN, the number of battles in a data base (601 for the HERO/CAA data base). The FORTRAN format of SEQ is INTEGER.

F-3. CODING SCHEME FOR HERO TABLE NUMBER 1

WAR Name of the war of which the battle/engagement is a part. Ref. HERO Table 1, CHARACTER*44.

NAME Name of the battle or engagement. Ref. HERO Table 1, CHARACTER*44.

LOCN Name of the place where the battle occurred (usually a nation or other geographical region). Ref. HERO Table 1, CHARACTER*44.

CAMPGN Name of the campaign of which this battle/engagement is a part. Ref. HERO Table 1, CHARACTER*60.

DATE Date on which the battle began, in the form YYYYMMDD where YYYY is the year, MM is the month number, and DD is the number of the day of the month. DATE is positive for AD dates and negative for BC dates. Ref. HERO Table 1, INTEGER.

T Duration of the battle, in days; an integer. Use -1 if unknown. Ref. HERO Table 1, INTEGER.

WOF Width of front in kilometers. Use -1.0 if unknown. Ref. HERO Table 1, REAL.

NAMA Name of the attacker's force that fought the battle. Ref HERO Table 1, CHARACTER*60.

COA Name of the commander of the attacker's forces in the battle. Ref. HERO Table 1, CHARACTER*60.

NAMD Name of the defender's force that fought the battle. Ref. HERO Table 1, CHARACTER*60.

COD Name of the commander of the defender's forces in the battle. Ref. HERO Table 1, CHARACTER*60.

F-4. CODING SCHEME FOR HERO TABLE NUMBER 2

POSTD1 Defender's primary defensive posture, categorized and coded as:

Code	Abbrv	Description
WD	WDL	Withdrawal from action
DL	Del	Delaying action
HD	HD	Hasty defense
PD	PD	Prepared defense
FD	FD	Fortified defense
00	--	None of the above

Ref. HERO Table 2, CHARACTER*5.

POSTD2 Defender's secondary defensive posture category. See POSTD1 for categories and codes. Ref. HERO Table 2, CHARACTER*5.

TERRA1 Three-character, primary terrain descriptor, categorized and coded as follows.

First Character:

Code	Abbrv	Description
G	Rg	Rugged
R	R	Rolling
F	F	Flat
0	--	Other

Second Character:

Code	Abbrv	Description
W	W	Heavily wooded
M	M	Mixed
B	B	Bare
D	D	Desert
0	--	Other

Third Character:

Code	Abbrv	Description
U	U	Urban
M	M	Marsh or swamp
D	Du	Dune
O	--	Other

Ref. HERO Table 2, Character*5

TERRA2 Three-character, secondary terrain descriptor. See TERRA1 for categories and codes. Ref. HERO Table 2, CHARACTER*5.

WX1 First five-character weather, season, and climate descriptor, categorized and coded as follows.

First Character:

Code	Abbrv	Description
W	W	Wet (i.e., precipitation)
D	D	Dry (i.e., no precipitation)
O	--	Other

Second Character:

Code	Abbrv	Description
H	H	Heavy (precipitation)
L	L	Light (precipitation)
O	O	Overcast (no precipitation)
S	S	Sunny
O	--	Other

Third Character:

Code	Abbrv	Description
H	H	Hot (local weather)
T	T	Temperate (local weather)
C	C	Cold (local weather)
O	--	Other

Fourth Character:

Code	Abbrv	Description
W	W	Winter (season)
$	Sp	Spring (season)
S	S	Summer (season)
F	F	Fall (season)
0	--	Other

Fifth Character:

Code	Abbrv	Description
E	TR	Tropical (climatic zone)
D	D	Desert (climate type)
T	T	Temperate (climatic zone)
0	--	Other

Ref. HERO Table 2, CHARACTER*10.

WX2 Second five-character weather, season, and climate descriptor. See WX1 for coding scheme. Ref. HERO Table 2, CHARACTER*10.

WX3 Third five-character weather, season, and climate descriptor. See WX1 for coding scheme. Ref. HERO Table 2, CHARACTER*10.

SURPA Relative surprise achieved by the attacker, categorized and coded as follows:

Code	Abbrv	Description
3	Complete	Complete surprise achieved by attacker
2	Substantial	Substantial surprise achieved by attacker
1	Minor	Minor surprise achieved by attacker
0	N	Surprise not achieved by either side, or did not influence the battle's outcome
-1	Minor	Minor surprise achieved by defender

Code	Abbrv	Description
-2	Substantial	Substantial surprise achieved by defender
-3	Complete	Complete surprise achieved by defender

Ref. HERO Table 2, INTEGER.

AEROA Relative air superiority achieved by the attacker, categorized and coded as follows:

Code	Abbrv	Description
1	x/	Air superiority in favor of the attacker
0	--	Neither side had air superiority, or it did not influence the battle
-1	/x	Air superiority in favor of the defender

Ref. HERO Table 2, INTEGER.

NOTE: POSTD1, POSTD2, TERRA1, TERRA2, WX1, WX2, and WX3 are all left-justified in their respective fields.

F-5. CODING SCHEME FOR HERO TABLE NUMBER 3

XO Total personnel strength of the attacker (-1 if unknown). Ref. HERO Table 3, INTEGER.

YO Total personnel strength of the defender (-1 if unknown). Ref. HERO Table 3, INTEGER.

CAVA Number of mounted troops (cavalry, dragoons, and mounted infantry) for the attacker (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

CAVD Number of mounted troops (cavalry, dragoons, and mounted infantry) for the defender (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

TANKA Total number of armored tank-like vehicles for the attacker (includes tanks; armored, self-propelled tank guns; and armored assault guns) (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

TANKD Total number of armored tank-like vehicles for the defender (includes tanks; armored, self-propelled tank guns; and armored assault guns) (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

LTA Total number of light armored tank-like vehicles for the attacker (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

LTD Total number of light armored tank-like vehicles for the defender (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

MBTA Total number of main battle tanks for the attacker (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

MBTD Total number of main battle tanks for the defender (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

ARTYA Total number of artillery pieces for the attacker (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

ARTYD Total number of artillery pieces for the defender (0 if none present, -1 if unknown). Ref. HERO Table 3, INTEGER.

FLYA Total number of air sorties flown in support of the attacker (0 if none flown, -1 if unknown). Ref. HERO Table 3, INTEGER.

FLYD Total number of air sorties flown in support of the defender (0 if none flown, -1 if unknown). Ref. HERO Table 3, INTEGER.

CX Battle casualties to the attacker's personnel (0 if none, -2 if unknown). Ref. HERO Table 3, INTEGER.

CY Battle casualties to the defender's personnel (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CTANKA Number of the attacker's tanks and other AFVs destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CTANKD Number of the defender's tanks and other AFVs destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CARTYA Number of the attacker's artillery pieces that were destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CARTYD Number of the defender's artillery pieces that were destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CFLYA Number of the attacker's combat aircraft lost as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

CFLYD Number of the defender's combat aircraft lost as a result of enemy action (0 if none, -1 if unknown). Ref. HERO Table 3, INTEGER.

F-6. CODING SCHEME FOR HERO TABLE NUMBER 4

CEA Attacker's adjudged relative advantage in combat effectiveness, categorized and coded as shown in the table below. The marginal entries are as described in paragraphs E-3c and e. The values used in CAA's computerized data base are shown in the body of the table. Ref. HERO Table 4, INTEGER.

		ABBRV. FOR ATTACKER						
		$\underline{0}$	o		x	$\underline{X}$	N	C
ABBRV. FOR DEFENDER	$\underline{0}$	0	1	2	3	4		
	o	-1	0	1	2	3		
		-2	-1	0	1	2	0	0
	x	-3	-2	-1	0	1		
	$\underline{X}$	-4	-3	-2	-1	0		
	N			0			0	
	C			0				0

LEADA Attacker's adjudged relative advantage in leadership. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

TRNGA Attacker's adjudged relative advantage in training and experience. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

MORALA Attacker's adjudged relative advantage in morale. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

LOGSA Attacker's adjudged relative advantage in logistics. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

MOMNTA Attacker's adjudged relative advantage in momentum. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

INTELA Attacker's adjudged relative advantage in (military) intelligence. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

TECHA Attacker's adjudged relative advantage in technology. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

INITA Attacker's adjudged relative advantage in initiative. See CEA for coding scheme. Ref. HERO Table 4, INTEGER.

F-7. CODING SCHEME FOR HERO TABLE NUMBER 5

WINA Attacker's adjudged relative level of victory, categorized and coded as follows:

Code	Abbrv	Description
1	x/	Attacker adjudged victorious
0	x/x	Drawn battle, neither side clearly victorious
-1	/x	Defender adjudged victorious

Ref. HERO Table 5, INTEGER.

KPDA Attacker's average rate of advance, in kilometers, per day. Use positive values for attacker's advance, negative values for defender's advance, and zero values for no (or negligible) advance. The value -9999 is used if the average rate of advance is unknown. Ref. HERO Table 5, REAL.

ACHA Attacker's adjudged mission accomplishment rating on a scale of 0 (mission not accomplished) to 10 (mission fully accomplished). The value -1 is used if the rating is unknown. Ref. HERO Table 5, INTEGER.

ACHD Defender's adjudged mission accomplishment rating on a scale of 0 (mission not accomplished) to 10 (mission fully accomplished). The value -1 is used if the rating is unknown. Ref. HERO Table 5, INTEGER.

F-8. CODING SCHEME FOR HERO TABLE 6

QUALA Attacker's adjudged relative force quality. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

RESA Attacker's adjudged relative skill in use of reserves. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

MOBILA Attacker's adjudged relative mobility superiority. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

AIRA Attacker's adjudged relative air superiority. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

FPREPA Attacker's adjudged relative force preponderance. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

WXA Attacker's adjudged relative weather advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

TERRA Attacker's adjudged relative terrain/roads advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

LEADAA Attacker's adjudged relative leadership advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

PLANA Attacker's adjudged relative planning effectiveness. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

SURPAA Attacker's adjudged relative surprise advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

MANA Attacker's adjudged relative maneuver advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

LOGSAA Attacker's adjudged relative logistics advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

FORTSA Attacker's adjudged relative fortification advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

DEEPA Attacker's adjudged relative depth advantage. Coded like CEA, see paragraph F-6. Ref. HERO Table 6, INTEGER.

F-9. CODING SCHEME FOR HERO TABLE NUMBER 7

PRIA1 Attacker's primary tactical scheme, part 1, categorized and coded as follows:

Code	Abbrv	Description
FF	F	Frontal attack
EE	E	Single envelopment
DE	EE	Double envelopment
FE	FE	Feint, demonstration, or holding attack
DD	D	Defensive plan
DO	D/O	Defensive/offensive plan
LF	(LF)	Left flank
RF	(RF)	Right flank
LR	(LR)	Left rear
RR	(RR)	Right rear
PP	P	Penetration
RC	RivC	River crossing
OO	--	None of the above

Ref. HERO Table 7, CHARACTER*4.

PRIA2 Attacker's primary tactical scheme, part 2. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

PRIA3 Attacker's primary tactical scheme, part 3. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECA1 Attacker's secondary tactical scheme, part 1. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECA2 Attacker's secondary tactical scheme, part 2. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECA3 Attacker's secondary tactical scheme, part 3. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

RESOA1 Attacker's resolution/outcome, part 1, categorized and coded as follows:

Code	Abbrv	Description
AA	A	Annihilated
PS	Ps	Pursued
WL	WDL	Withdrew with serious loss
WD	WD	Withdrew
BB	B	Breakthrough
PP	P	Penetration
RR	R	Repulse
SS	S	Stalemate
00	--	None of the above

Ref. HERO Table 7, CHARACTER*4.

RESOA2 Attacker's resolution/outcome, part 2. See RESOA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

RESOA3 Attacker's resolution/outcome, part 3. See RESOA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

PRID1 Defender's primary tactical scheme, part 1. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

PRID2 Defender's primary tactical scheme, part 2. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

PRID3 Defender's primary tactical scheme, part 3. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECD1 Defender's secondary tactical scheme, part 1. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECD2 Defender's secondary tactical scheme, part 2. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

SECD3 Defender's secondary tactical scheme, part 3. See PRIA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

RESOD1 Defender's resolution/outcome, part 1. See RESOA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

RESOD2 Defender's resolution/outcome, part 2. See RESOA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

RESOD3 Defender's resolution/outcome, part 3. See RESOA1 for coding scheme. Ref. HERO Table 7, CHARACTER*4.

WGT Relative adjudged rating of the accuracy/validity of the data for this battle. Not used in this data base. All battles assigned a weight rating of M (moderate accuracy/validity). CHARACTER*4.

NOTE: PRIA1, PRIA2, PRIA3, SECA1, SECA2, SECA3, RESOA1, RESOA2, RESOA3, PRID1, PRID2, PRID3, SECD1, SECD2, SECD3, RESOD1, RESOD2, and RESOD3 are all right-justified in their respective fields.

APPENDIX G

BATTLE DATA FILE FORMATS

G-1. BATTLE SEQUENCE NUMBER. This is an index number, and so is not tabulated with the other data values.

G-2. FORMAT FOR FILE 03TABLE1

a. Description. This file is based on HERO's Table 1 and is arranged in ascending order by battle sequence number. It contains four records for each battle sequence number.

b. Variables and Formats

Record no.	Format no.	Format	Variables
1	511	(3A44)	WAR, NAME, LOCN
2	512	(A60,I10,I5,F10.1)	CAMPGN, DATE, T, WOF
3	513	(2A60)	NAMA, COA
4	513	(2A60)	NAMD, COD

c. Tributary Files

Filename	Format	Variables
03WAR.	(A60)	WAR
03NAME.	(A60)	NAME
03DATE.	(I10)	DATE
03T.	(I5)	T
03WOF	(F10.1)	WOF
03LOCAP.	(A59,A60)	LOCN, CAMPGN
03ATTID.	(A59,A60)	NAMA, COA
03DEFID.	(A59,A60)	NAMD, COD

G-3. FORMAT FOR FILE 03TABLE2

a. Description. This file is based on HERO's Table 2 and is arranged in ascending order by battle sequence number. It contains one record for each battle sequence number.

b. Variables and Formats

Record no.	Format no.	Format	Variables
1	521	(4A5,3A10,2I5)	POSTD1, POSTD2, TERRA1, TERRA2, WX1, WX2, WX3, SURPA, AEROA

c. **Tributary Files**

Filename	Format	Variables
03TABLE2A.	(4A5,A10,2I5)	POSTD1, POSTD2, TERRA1, TERRA2, WX1, SURPA, AEROA
03TABLE2B.	(4A10)	WX2, WX3

G-4. FORMAT FOR FILE 03TABLE3

a. **Description.** This file is based on HERO's Table 3 and is arranged in ascending order by battle sequence number. It contains two records for each battle sequence number.

b. **Variables and Formats**

Record no.	Format no.	Format	Variables
1	531	(11I10)	XO, CAVA, TANKA, LTA, MBTA, ARTYA, FLYA, CX, CTANKA, CARTYA, CFLYA
2	531	(11I10)	YO, CAVD, TANKD, LTD, MBTD, ARTYD, FLYD, CY, CTANKD, CARTYD, CFLYD

c. **Tributary Files**

Filename	Format	Variables
03XO.	(I10)	XO
03YO.	(I10)	YO
03CX.	(I10)	CX
03CY.	(I10)	CY
03CAV.	(2I10)	CAVA,CAVD
03TANK.	(2I10)	TANKA,TANKD
03LT.	(2I10)	LTA,LTD
03MBT.	(2I10)	MTBA,MBTD
03ARTY.	(2I10)	ARTYA,ARTYD
03FLY.	(2I10)	FLYA,FLYD
03CTANK.	(2I10)	CTANKA,CTANKD
03CARTY.	(2I10)	CARTYA,CARTYD
03CFLY.	(2I10)	CFLYA,CFLYD

G-5. FORMAT FOR FILE 03TABLE4

a. **Description.** This file is based on HERO's Table 4 and is arranged in ascending order by battle sequence number. It contains one record for each battle sequence number.

b. Variables and Formats

Record no.	Format no.	Format	Variables
1	541	(9I5)	CEA, LEADA, TRNGA, MORALA, LOGSA, MOMNTA, INTELA, TECHA, INITA

c. Tributary Files. None.

G-6. FORMAT FOR FILE 03TABLE5

a. Description. This file is based on HERO's Table 5 and is arranged in ascending order by battle sequence number. It contains one record for each battle sequence number.

b. Variables and Formats

Record no.	Format no.	Format	Variables
1	551	(I5,F10.1,2I5)	WINA, KPDA, ACHA, ACHD

c. Tributary Files

Filename	Format	Variables
03WINA.	(I5)	WINA
03KPDA.	(F10.1)	KPDA
03ACHA.	(I5)	ACHA
03ACHD.	(I5)	ACHD

G-7. FORMAT FOR FILE 03TABLE6

a. Description. This file is based on HERO's Table 6 and is arranged in ascending order by battle sequence number. It contains one record for each battle sequence number.

b. Variables and Formats

Record no.	Format no.	Format	Variables
1	561	(14I5)	QUALA, RESA, MOBILA, AIRA, FPREPA, WXA, TERRA, LEADAA, PLANA, SURPAA, MANA, LOGSAA, FORTSA, DEEPA

c. **Tributary Files**

Filename	Format	Variables
03TABLE6A.	(4I5)	QUALA, RESA, MOBILA, AIRA
03TABLE6B.	(5I5)	FPREPA, WXA, TERRA, LEADAA, PLANA
03TABLE6C.	(5I5)	SURPAA, MANA, LOGSAA, FORTSA, DEEPA

G-8. FORMAT FOR FILE 03TABLE7

a. **Description.** This file is based on HERO's Table 7 and is arranged in ascending order by battle sequence number. It contains three records for each battle sequence number.

b. **Variables and Formats**

Record no.	Format no.	Format	Variables
1	571	(9A4)	PRIA1,PRIA2, PRIA3, SECA1, SECA2, SECA3, RESOA1, RESOA2, RESOA3
2	571	(9A4)	PRID1, PRID2, PRID3, SECD1, SECD2, SECD3, RESOD1, RESOD2, RESOD3
3	572	(A4)	WGT

c. **Tributary Files**

Filename	Format	Variables
03TABLE7A.	(9A4)	PRIA1, PRIA2, PRIA3, SECA1, SECA2, SECA3, RESOA1, RESOA2, RESOA3
03TABLE7B.	(9A4)	PRID1, PRID2, PRID3, SECD1, SECD2, SECD3, RESOD1, RESOD2, RESOD3

APPENDIX H

INDEX OF BATTLES AND ENGAGEMENTS IN THE COMPUTERIZED DATA BASE

ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
1	2	NIEUPORT	1600	JUL	2	NIEUPORT 1600
2	2	WHITE MOUNTAIN	1620	NOV	8	BOHEMIA 1620
3	2	WIMPFEN	1622	MAY	6	PALATINATE 1622
4	2	DESSAU BRIDGE	1626	APR	25	DANISH INVASION OF GERMANY 1625-26
5	2	LUTTER	1626	AUG	27	DANISH INVASION OF GERMANY 1625-26
6	2	BREITENFELD I	1631	SEP	17	LEIPZIG 1631
7	2	THE LECH	1632	APR	15	BAVARIA 1632
8	2	ALTE VESTE	1632	SEP	3	NUREMBERG 1632
9	2	LUETZEN	1632	NOV	16	SAXONY 1632
10	2	NORDLINGEN I	1634	SEP	6	BAVARIA 1634
11	2	WITTSTOCK	1636	OCT	4	E GERMANY 1636
12	2	BREITENFELD II	1642	NOV	2	SAXONY 1642
13	2	ROCROI	1643	MAY	19	NE FRANCE 1643
14	2	TUTTLINGEN	1643	NOV	24	SWABIA 1643
15	2	FREIBURG	1644	AUG	3	SWABIA 1644
16	2	JANKAU	1645	MAR	6	BOHEMIA 1645
17	2	MERGENTHEIM	1645	MAY	2	BAVARIA 1645
18	2	ALLERHEIM (NORDLINGEN II)	1645	AUG	3	BAVARIA 1645
19	2	LENS	1648	AUG	10	NE FRANCE 1648
20	2	EDGEHILL	1642	OCT	23	EDGEHILL
21	2	MARSTON MOOR	1644	JUL	2	YORK
22	2	TIPPERMUIR	1644	SEP	1	ABERDEEN
23	2	KILSYTH	1644	AUG	15	KILSYTH
24	2	NEWBURY II	1644	SEP	27	NEWBURY II
25	2	NASEBY	1645	JUN	14	NASEBY
26	2	PRESTON	1648	AUG	17	PRESTON
27	2	DUNBAR	1650	SEP	3	DUNBAR
28	2	WORCESTER	1651	SEP	3	WORCESTER
29	2	ST. ANTOINE	1652	JUL	5	THE FRONDE
30	2	THE DUNES	1658	JUN	14	DUNKIRK 1658
31	2	THE RAAB	1664	AUG	1	HUNGARY 1664
32	2	VIENNA	1683	SEP	12	AUSTRIA 1683
33	2	CHOCIM II	1673	NOV	11	CHOCIM
34	2	SINSHEIM	1674	JUN	16	RHINELAND 1674
35	2	SENEF	1674	AUG	11	SPANISH NETHERLANDS 1674
36	2	ENZHEIM	1674	OCT	4	RHINELAND 1674
37	2	TURCKHEIM	1675	JAN	5	RHINELAND 1675
38	2	FEHRBELLIN	1675	JUN	28	BRANDENBURG 1675
39	2	SEDGEMOOR	1685	JUL	6	SEDGEMOOR
40	2	KILLIECRANKIE	1689	JUL	27	KILLIECRANKIE
41	2	WALCOURT	1689	AUG	25	FLANDERS 1689
42	2	FLEURUS	1690	JUL	1	FLEURUS 1690
43	2	THE BOYNE	1690	JUL	11	BOYNE
44	2	AUGHRIM	1691	JUL	22	AUGHRIM
45	2	STEENKERKE	1692	AUG	3	FLANDERS 1692
46	2	NEERWINDEN (LANDEN)	1693	JUL	29	FLANDERS 1693
47	2	MARSAGLIA	1693	OCT	4	PIEDMONT 1693
48	2	ZENTA	1697	SEP	11	HUNGARY 1697
49	2	POLTAVA	1709	JUN	28	POLTAVA
50	2	BLENHEIM	1704	AUG	13	BLENHEIM
51	2	RAMILLIES	1706	MAY	23	RAMILLIES
52	2	OUDENARDE	1708	JUL	11	OUDENARDE
53	2	MALPLAQUET	1709	SEP	11	MALPLAQUET
54	2	PETERWARDEIN	1716	AUG	5	HUNGARY 1716
55	2	MOLLWITZ	1741	APR	10	SILESIAN
56	2	CHOTUSITZ	1742	MAY	17	BOHEMIA 1742
57	2	DETTINGEN	1743	JUN	27	DETTINGEN
58	2	FONTENOY	1745	MAY	11	FONTENOY
59	2	HOHENFRIEDBERG	1745	JUN	4	HOHENFRIEDBERG
60	2	SOOR	1745	SEP	30	SOOR
61	2	KESSELDORF	1745	DEC	14	ELBE
62	2	PRESTONPANS	1745	SEP	21	PRESTONPANS
63	2	CULLODEN	1746	APR	16	CULLODEN
64	2	LOBOSITZ	1756	OCT	1	PIRNA-LOBOSITZ
65	2	PRAGUE	1757	MAY	6	BOHEMIA 1757
66	2	PLASSEY	1757	JUN	23	W BENGAL
67	2	KOLIN	1757	JUN	18	BOHEMIA 1757
68	2	HASTENBECK	1757	JUL	26	HASTENBECK
69	2	ROSSBACH	1757	NOV	5	ROSSBACH
70	2	LEUTHEN	1757	DEC	5	LEUTHEN

71	2	CREFELD	1758	JUN	23	RHINELAND 1758
72	2	ZORNDORF	1758	AUG	25	ZORNDORF
73	2	HOCHKIRCH	1758	OCT	14	HOCHKIRK
74	2	BERGEN	1759	APR	13	BERGEN
75	2	MINDEN	1759	AUG	1	MINDEN
76	2	KUNERSDORF	1759	AUG	12	KUNERSDORF
77	2	PLAINS OF ABRAHAM (QUEBEC)	1759	SEP	13	QUEBEC
78	2	MAXEN	1759	NOV	21	MAXEN
79	2	WARBURG	1760	JUL	31	HONOVER 1760
80	2	LIEGNITZ	1760	AUG	15	SILESIA 1760
81	2	TORGAU	1760	NOV	3	SILESIA 1760
82	2	BUNKER HILL	1775	JUN	17	SIEGE OF BOSTON
83	2	QUEBEC	1775	DEC	31	CANADA INVASION 1775-76
84	2	WHITE PLAINS	1776	OCT	28	NEW YORK 1776
85	2	TRENTON	1776	DEC	26	NEW JERSEY 1776-77
86	2	PRINCETON	1777	JAN	3	NEW JERSEY 1776-77
87	2	FREEMAN'S FARM	1777	SEP	19	SARATOGA
88	2	GERMANTOWN	1777	OCT	4	PHILADELPHIA 1777-78
89	2	BEMIS HEIGHTS	1777	OCT	7	SARATOGA
90	2	MONMOUTH COURT HOUSE	1778	JUN	28	NEW JERSEY 1778
91	2	CAMDEN	1780	AUG	16	CAMDEN
92	2	COWPENS	1781	JAN	17	SOUTHERN 1780-81
93	2	GUILFORD COURT HOUSE	1781	MAR	15	SOUTHERN 1780-81
94	2	HOBKIRK'S HILL	1781	APR	25	SOUTHERN 1780-81
95	2	EUTAW SPRINGS	1781	SEP	8	SOUTHERN 1780-81
96	2	VALMY	1792	SEP	20	FRANCE 1792
97	2	JEMAPPES	1792	NOV	6	FLANDERS 1792
98	2	NEERWINDEN	1793	MAR	18	FLANDERS 1793
99	2	HONDSCHOOTE	1793	SEP	6	FLANDERS 1793
100	2	WATTIGNIES	1793	OCT	15	FLANDERS 1793
101	2	FLEURUS	1794	JUN	26	FLANDERS 1794
102	2	LODI	1796	MAY	10	ITALY 1796
103	2	CASTIGLIONE	1796	AUG	5	ITALY 1796
104	2	NERESHEIM	1796	AUG	11	GERMANY 1796
105	2	WURZBURG	1796	SEP	3	GERMANY 1796
106	2	ARCOLA	1796	NOV	15	ITALY 1796
107	2	RIVOLI	1797	JAN	14	ITALY 1797
108	2	PYRAMIDS	1798	JUL	21	EGYPT
109	2	STOCKACH I	1799	MAR	25	GERMANY 1799
110	2	MOUNT TABOR	1799	APR	16	EGYPT (PALESTINE)
111	2	ZURICH I	1799	JUN	4	SWITZERLAND 1799
112	2	NOVI	1799	AUG	15	ITALY 1799
113	2	ZURICH III	1799	SEP	24	SWITZERLAND 1799
114	2	MOSKIRCH	1800	MAY	5	GERMANY 1800
115	2	MARENGO	1800	JUN	14	ITALY 1800
116	2	HOHENLINDEN	1800	DEC	3	GERMANY 1800

ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
117	3	AUSTERLITZ	1805	DEC	2	1805
118	3	JENA	1806	OCT	14	JENA
119	3	AUERSTADT	1806	OCT	14	JENA
120	3	EYLAU	1807	FEB	8	POLAND 1807
121	3	FRIEDLAND	1807	JUN	14	POLAND 1807
122	3	VIMIERO	1808	AUG	21	PENINSULAR 1808
123	3	CORUNNA	1809	JAN	16	PENINSULAR 1809
124	3	ECKMUEHL	1809	APR	22	WAGRAM
125	3	ASPERN-ESSLING	1809	MAY	21	WAGRAM
126	3	THE RAAB	1809	JUN	14	WAGRAM (ITALY)
127	3	WAGRAM	1809	JUL	5	WAGRAM
128	3	TALAVERA	1809	JUL	28	PENINSULAR 1809
129	3	BUSSACO	1810	SEP	27	PENINSULAR 1810
130	3	FUENTES DE ONORO	1811	MAY	5	PENINSULAR 1811
131	3	ALBUERA	1811	MAY	16	PENINSULAR 1811
132	3	SALAMANCA	1812	JUL	22	PENINSULAR 1812
133	3	VITTORIA	1813	JUN	21	PENINSULAR 1813
134	3	BORODINO	1812	SEP	7	RUSSIA 1812
135	3	LUETZEN	1813	MAY	2	LEIPZIG 1813
136	3	BAUTZEN	1813	MAY	20	LEIPZIG 1813
137	3	DRESDEN	1813	AUG	26	LEIPZIG 1813
138	3	LEIPZIG	1813	OCT	16	LEIPZIG 1813
139	3	HANAU	1813	OCT	30	LEIPZIG 1813
140	3	LA ROTHIERE	1814	FEB	1	DEFENSE OF FRANCE
141	3	LAON	1814	MAR	9	DEFENSE OF FRANCE
142	3	ARCIS-SUR-AUBE	1814	MAR	20	DEFENSE OF FRANCE

143	3	LIGNY	1815	JUN	16	THE HUNDRED DAYS
144	3	QUATRE BRAS	1815	JUN	16	THE HUNDRED DAYS
145	3	WATERLOO	1815	JUN	18	THE HUNDRED DAYS
146	3	THE THAMES	1813	OCT	5	NORTHWESTERN
147	3	CHIPPEWA	1814	JUL	5	NORTHERN
148	3	LUNDY'S LANE	1814	JUL	25	NORTHERN
149	3	NEW ORLEANS	1815	JAN	8	NEW ORLEANS
150	3	BOYACA	1819	AUG	7	BOYACA
151	3	CARABORO	1821	JUN	25	CARAROBO
152	3	BOMBONA	1822	APR	7	BOMBONA
153	3	PICHINCHA	1822	MAY	24	PICHINCHA
154	3	JUNIN	1824	AUG	6	JUNIN
155	3	AYACUCHO	1824	DEC	9	AYACUCHO
156	3	SAN JACINTO	1836	APR	21	TEXAS 1836
157	3	PALO ALTO	1846	MAY	8	NORTHERN
158	3	RESACA DE LA PALMA	1846	MAY	9	NORTHERN
159	3	BUENA VISTA	1847	FEB	22	NORTHERN
160	3	CERRO GORDO	1847	APR	17	CENTRAL MEXICAN
161	3	CONTRERAS	1847	AUG	20	CENTRAL MEXICAN
162	3	CHURUBUSCO	1847	AUG	20	CENTRAL MEXICAN
163	3	MOLINO DEL REY	1847	SEP	8	CENTRAL MEXICAN
164	3	CHAPULTEPEC	1847	SEP	13	CENTRAL MEXICAN
165	3	THE ALMA	1854	SEP	20	SEBASTOPOL
166	3	INKERMAN	1854	NOV	5	SEBASTOPOL
167	3	MAGENTA	1859	JUN	4	LOMBARDY 1859
168	3	SOLFERINO	1859	JUN	24	LOMBARDY 1859
169	3	SADOWA	1866	JUL	3	BOHEMIA 1866
170	3	CUSTOZZA II	1866	JUN	24	VENETIA 1866
171	3	FIRST BULL RUN (MANASSAS)	1861	JUL	21	FIRST BULL RUN
172	3	WILSON'S CREEK	1861	AUG	10	MISSOURI 1861
173	3	BELMONT	1861	NOV	7	MISSOURI 1861
174	3	MILL SPRINGS	1862	JAN	19	KENTUCKY 1862
175	3	FORT DONELSON	1862	FEB	15	HENRY AND DONELSON
176	3	PEA RIDGE	1862	MAR	7	MISSOURI 1862
177	3	KERNSTOWN	1862	MAR	23	VALLEY 1862
178	3	SHILOH	1862	APR	6	TENNESSEE 1862
179	3	FRONT ROYAL	1862	MAY	23	VALLEY 1862
180	3	FIRST WINCHESTER	1862	MAY	25	VALLEY 1862
181	3	CROSS KEYS	1862	JUN	8	VALLEY 1862
182	3	PORT REPUBLIC	1862	JUN	9	VALLEY 1862
183	3	SEVEN PINES (FAIR OAKS)	1862	MAY	31	PENINSULAR 1862
184	3	MECHANICSVILLE	1862	JUN	26	PENINSULAR 1862
185	3	GAINES'S MILL	1862	JUN	27	PENINSULAR 1862
186	3	GLENDALE-FRAYSER'S FARM	1862	JUN	29	PENINSULAR 1862
187	3	MALVERN HILL	1862	JUL	1	PENINSULAR 1862
188	3	CEDAR MOUNTAIN	1862	AUG	9	SECOND BULL RUN
189	3	SECOND BULL RUN (MANASSAS)	1862	AUG	29	SECOND BULL RUN
190	3	SOUTH MOUTAIN	1862	SEP	14	ANTIETAM
191	3	ANTIETAM	1862	SEP	17	ANTIETAM
192	3	CORINTH	1862	OCT	3	IUKA-CORINTH
193	3	PERRYVILLE	1862	OCT	8	PERRYVILLE
194	3	FREDERICKSBURG	1862	DEC	13	FREDRICKSBURG
195	3	MURFREESBORO	1862	DEC	31	STONES RIVER
196	3	CHANCELLORSVILLE	1863	MAY	1	CHANCELLORSVILLE
197	3	CHAMPION'S HILL	1863	MAY	16	VICKSBURG
198	3	BRANDY STATION	1863	JUN	9	GETTYSBURG
199	3	GETTYSBURG	1863	JUL	1	GETTYSBURG
200	3	CHICKAMAUGA	1863	SEP	19	CHICKAMAUGA
201	3	CHATTANOOGA	1863	NOV	24	CHATTANOOGA
202	3	THE WILDERNESS	1864	MAY	5	WILDERNESS
203	3	SPOTSYLVANIA	1864	MAY	8	SPOTSYLVANIA
204	3	NEW MARKET	1864	MAY	15	SHENANDOAH VALLEY 1864
205	3	COLD HARBOR	1864	JUN	3	WILDERNESS-SPOTSYLVANIA-COLD HARBOR
206	3	KENESAW MOUNTAIN	1864	JUN	27	ATLANTA
207	3	PEACHTREE CREEK	1864	JUL	20	ATLANTA
208	3	ATLANTA	1864	JUL	22	ATLANTA
209	3	PETERSBURG	1864	JUN	15	PETERSBURG
210	3	GLOBE TAVERN	1864	AUG	18	SIEGE OF PETERSBURG
211	3	OPEQUON CREEK	1864	SEP	19	SHERIDAN'S VALLEY
212	3	CEDAR CREEK	1864	OCT	19	SHERIDAN'S VALLEY
213	3	FRANKLIN	1864	NOV	30	FRANKLIN AND NASHVILLE
214	3	NASHVILLE	1864	DEC	15	FRANKLIN AND NASHVILLE

215	3	BENTONVILLE	1865	MAR	19	THE CAROLINAS
216	3	DINWIDDIE COURTHOUSE	1865	MAR	29	PETERSBURG
217	3	FIVE FORKS	1865	APR	1	PETERSBURG
218	3	SELMA	1865	APR	2	SELMA
219	3	SAYLOR'S CREEK	1865	APR	6	APPOMATTOX
220	3	WEISSENBURG	1870	AUG	4	METZ
221	3	FROESCHWILLER	1870	AUG	6	METZ
222	3	SPICHERN	1870	AUG	6	METZ
223	3	MARS LA TOUR	1870	AUG	16	METZ
224	3	GRAVELOTTE-ST. PRIVAT	1870	AUG	18	METZ
225	3	SEDAN	1870	SEP	1	SEDAN
226	3	COULMIERS	1870	NOV	9	ORLEANS
227	3	ORLEANS	1870	DEC	2	ORLEANS
228	3	LE MANS	1871	JAN	11	LOIRE
229	3	BELFORT	1871	JAN	15	BELFORT
230	3	ISANDHLWANA	1879	JAN	22	ZULULAND 1879
231	3	ULUNDI	1879	JUL	4	ZULULAND 1879
232	3	MAJUBA HILL	1881	FEB	27	SOUTH AFRICA 1881
233	3	TEL EL-KEBIR	1882	SEP	13	EGYPT 1882
234	3	OMDURMAN	1898	SEP	2	THE SUDAN 1898
235	3	ADOWA	1896	MAR	1	ETHIOPIA 1896
236	3	MODDER RIVER	1899	NOV	28	KIMBERLY
237	3	MAGERSFONTEIN	1899	DEC	11	KIMBERLY
238	3	COLENSO	1899	DEC	15	LADYSMITH
239	3	SPION KOP	1900	JAN	24	LADYSMITH
240	3	PAARDEBURG	1900	FEB	18	LADYSMITH
241	3	SAN JUAN HILL	1898	JUL	1	SANTIAGO

ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
242	4	THE YALU	1904	APR	30	YALU 1904
243	4	TELISSU	1904	JUN	14	MANCHURIAN 1904
244	4	LIAOYANG	1904	AUG	25	MANCHURIAN 1904
245	4	THE SHA-HO	1904	OCT	5	MANCHURIAN 1904
246	4	SANDEPU	1905	JAN	26	MANCHURIAN 1905
247	4	MUKDEN	1905	FEB	21	MANCHURIAN 1905
248	4	KUMANOVO	1912	OCT	23	MACEDONIA 1912
249	4	LULE BURGAS	1912	OCT	28	THRACE 1912
250	4	PRELIP	1912	NOV	1	MACEDONIA 1912
251	4	MONASTIR	1912	NOV	5	MACEDONIA 1912
252	4	ADRIANOPLE	1913	MAR	23	THRACE 1913
253	4	WARSAW	1920	AUG	14	POLISH COUNTEROFFENSIVE
254	4	THE NIEMAN	1920	SEP	23	POLISH OFFENSIVE SEP-OCT 1920
255	4	GUADALAJARA-BRIHUEGA	1937	MAR	11	MADRID 1937
256	4	CHANGKUFENG-SHACHAOFENG	1938	JUL	30	CHANGKUFENG
257	4	HILL 52-SHACHAOFENG	1938	AUG	2	CHANGKUFENG
258	4	CHANGKUFENG-HILL 52	1938	AUG	6	CHANGKUFENG
259	4	NOMONHAN-OPENING ENGAGEMENT	1939	MAY	23	NOMONHAN
260	4	NOMONHAN-SOVIET COUNTEROFFENSIVE	1939	AUG	20	NOMONHAN
261	4	SUOMUSSALMI	1939	DEC	11	FINLAND 1939-40
262	4	ALSACE-LORRAINE I	1914	AUG	15	THE FRONTIERS
263	4	ALSACE-LORRAINE II	1914	AUG	20	THE FRONTIERS
264	4	THE ARDENNES	1914	AUG	22	THE FRONTIERS
265	4	THE SAMBRE	1914	AUG	22	THE FRONTIERS
266	4	MONS	1914	AUG	23	THE FRONTIERS
267	4	LE CATEAU	1914	AUG	26	ADVANCE TO THE MARNE
268	4	GUISE	1914	AUG	29	ADVANCE TO THE MARNE
269	4	HEIGHTS OF NANCY	1914	SEP	3	THE MARNE 1914
270	4	OURCQ I	1914	SEP	5	THE MARNE 1914
271	4	OURCQ II	1914	SEP	6	THE MARNE 1914
272	4	PETIT MORIN	1914	SEP	6	THE MARNE 1914
273	4	TWO MORINS	1914	SEP	6	THE MARNE 1914
274	4	MARSHES OF ST.GOND	1914	SEP	6	THE MARNE 1914
275	4	VITRY LE FRANCOIS	1914	SEP	6	THE MARNE 1914
276	4	GAP OF REVIGNY	1914	SEP	6	THE MARNE 1914
277	4	THE AISNE	1914	SEP	13	RETREAT FROM THE MARNE
278	4	STALLUPONEN	1914	AUG	17	E PRUSSIA 1914
279	4	GUMBINNEN	1914	AUG	20	E PRUSSIA 1914
280	4	TANNENBERG	1914	AUG	26	E PRUSSIA 1914
281	4	MASURIAN LAKES	1914	SEP	9	E PRUSSIA 1914
282	4	KRASNIK	1914	AUG	23	GALICIA 1914
283	4	KOMAROV	1914	AUG	26	GALICIA 1914
284	4	GNILA LIPA	1914	AUG	26	GALICIA 1914

285	4	RAVA RUSSKA	1914	SEP	3	GALICIA 1914
266	4	LODZ	1914	NOV	12	W POLAND 1914
287	4	THE JADAR	1914	AUG	12	SERBIA 1914
288	4	THE KOLUBRA	1914	DEC	3	SERBIA 1914
289	4	EASTERN CHAMPAGNE	1915	FEB	15	NOYON SALIENT
290	4	NEUVE CHAPELLE	1915	MAR	10	NOYON SALIENT
291	4	YPRES II	1915	APR	22	NOYON SALIENT
292	4	FESTUBERT	1915	MAY	16	NOYON SALIENT
293	4	LOOS	1915	SEP	25	NOYON SALIENT
294	4	WINTER BATTLE	1915	FEB	7	EASTERN FRONT 1915
295	4	GOLICE-TARNOW (OPENING PHASE)	1915	MAY	2	EASTERN FRONT 1915
296	4	FIRST ISONZO	1915	JUN	23	ISONZO FRONT 1915
297	4	SECOND ISONZO	1915	JUL	18	ISONZO FRONT 1915
298	4	THIRD ISONZO	1915	OCT	18	ISONZO FRONT 1915
299	4	FOURTH ISONZO	1915	NOV	10	ISONZO FRONT 1915
300	4	FIRST DARDANELLES LANDING	1915	APR	25	GALLIPOLI
301	4	SUVLA BAY	1915	AUG	7	GALLIPOLI
302	4	KUT-EL-AMARA	1915	SEP	27	MESOPOTAMIA 1915
303	4	CTESIPHON	1915	NOV	22	MESOPOTAMIA 1915
304	4	FIRST SOMME	1916	JUL	1	WESTERN FRONT 1916
305	4	SOMME-FOURTH ARMY ATTACK	1916	JUL	1	SOMME
306	4	SOMME-OVILLERS	1916	JUL	1	SOMME
307	4	SOMME-BAZENTIN RIDGE	1916	JUL	14	SOMME
308	4	SOMME-FLERS-COURCELETTE	1916	SEP	15	SOMME
309	4	CAUCASUS WINTER OFFENSIVE	1916	JAN	10	CAUCASUS 1916
310	4	LAKE NAROTCH	1916	MAR	18	EASTERN FRONT 1916
311	4	1916 BRUSILOV OFFENSIVE	1916	JUN	4	EASTERN FRONT 1916
312	4	FIFTH ISONZO	1916	MAR	11	ISONZO FRONT 1916
313	4	ASIAGO	1916	MAY	15	AUSTRIAN TRENTINO OFFENSIVE
314	4	TRENTINO COUNTER-OFFENSIVE	1916	JUN	16	AUSTRIAN TRENTINO OFFENSIVE
315	4	SIXTH ISONZO (GORIZIA)	1916	AUG	6	ISONZO FRONT 1916
316	4	ARRAS	1917	APR	9	WESTERN FRONT 1917
317	4	AISNE II	1917	APR	16	WESTERN FRONT 1917
318	4	MESSINES	1917	JUN	7	FLANDERS 1917
319	4	YPRES III	1917	JUL	31	FLANDERS 1917
320	4	CAMBRAI I	1917	NOV	20	WESTERN FRONT 1917
321	4	CAMBRAI II	1917	NOV	30	WESTERN FRONT 1917
322	4	TENTH ISONZO	1917	MAY	12	ISONZO FRONT 1917
323	4	ELEVENTH ISONZO	1917	AUG	18	ISONZO FRONT 1917
324	4	CAPORETTO	1917	OCT	24	ISONZO FRONT 1917
325	4	TIGRIS CROSSING	1917	FEB	22	MESOPOTAMIA 1917
326	4	GAZA I	1917	MAR	26	PALESTINE 1917
327	4	GAZA II	1917	APR	17	PALESTINE 1917
328	4	GAZA III	1917	OCT	31	PALESTINE 1917
329	4	JUNCTION STATION	1917	NOV	13	PALESTINE 1917
330	4	SECOND SOMME-PHASE I (SOMME-PERONNE)	1918	MAR	21	GERMAN SPRING OFFENSIVES
331	4	SECOND SOMME-PHASE II (SOMME-MONTDIDIER)	1918	MAR	27	GERMAN SPRING OFFENSIVES
332	4	LYS	1918	APR	9	GERMAN SPRING OFFENSIVES
333	4	YVONNE & ODETTE POSITIONS (SECTOR TOULON)	1918	APR	13	VERDUN SECTOR
334	4	CHEMIN-DES-DAMES	1918	MAY	27	GERMAN SPRING OFFENSIVES
335	4	CANTIGNY	1918	MAY	28	GERMAN SPRING OFFENSIVES
336	4	BELLEAU WOOD	1918	JUN	6	BELLEAU WOOD
337	4	HILL 142	1918	JUN	6	BELLEAU WOOD
338	4	WEST WOOD I	1918	JUN	6	BELLEAU WOOD
339	4	BOURESCHES I	1918	JUN	6	BELLEAU WOOD
340	4	HILL 192	1918	JUN	6	BELLEAU WOOD
341	4	WEST WOOD II	1918	JUN	11	BELLEAU WOOD
342	4	NORTH WOOD I (HUNTING LODGE)	1918	JUN	12	BELLEAU WOOD
343	4	BOURESCHES II	1918	JUN	13	BELLEAU WOOD
344	4	NORTH WOOD II	1918	JUN	21	BELLEAU WOOD
345	4	NORTH WOOD III	1918	JUN	23	BELLEAU WOOD
346	4	NORTH WOOD IV	1918	JUN	25	BELLEAU WOOD
347	4	VAUX	1918	JUL	1	BELLEAU WOOD
348	4	LA ROCHE WOOD EAST	1918	JUL	1	BELLEAU WOOD
349	4	LA ROCHE WOOD WEST	1918	JUL	1	BELLEAU WOOD
350	4	NOYON-MONTDIDIER	1918	JUN	9	BELLEAU WOOD
351	4	CHAMPAGNE-MARNE	1918	JUL	15	GERMAN SPRING OFFENSIVE 1918
352	4	AISNE-MARNE I	1918	JUL	18	AISNE-MARNE
353	4	MISSY AUX BOIS RAVINE	1916	JUL	18	SOISSONS
354	4	BREUIL	1918	JUL	18	SOISSONS
355	4	ST. AMAND FARM	1918	JUL	18	SOISSONS
356	4	BEAUREPAIRE FARM	1918	JUL	18	SOISSONS

357	4	CRAVANCON FERME-CHAUDUN	1918	JUL	18	SOISSONS
358	4	CHAUDUN	1918	JUL	18	SOISSONS
359	4	AISNE-MARNE II	1918	JUL	20	AISNE-MARNE
360	4	BERZY LE SEC	1918	JUL	21	AISNE-MARNE II
361	4	BUZANCY RIDGE	1918	JUL	21	AISNE-MARNE II
362	4	PICARDY 1918, I	1918	AUG	8	AMIENS OFFENSIVE
363	4	PICARDY 1918, II	1918	AUG	21	AMIENS OFFENSIVE
364	4	ST. MIHIEL	1918	SEP	12	ST. MIHIEL
365	4	LAHAYVILLE-BOIS DE LAMARCHE	1918	SEP	12	ST. MIHIEL
366	4	MEUSE-ARGONNE I	1918	SEP	26	MEUSE-ARGONNE
367	4	BLANC-MONT I	1918	OCT	3	MEUSE-ARGONNE (CHAMPAGNE)
368	4	MEDEAH FARM	1918	OCT	3	MEUSE-ARGONNE (CHAMPAGNE)
369	4	ESSEN HOOK	1918	OCT	3	MEUSE-ARGONNE (CHAMPAGNE)
370	4	BLANC MONT RIDGE	1918	OCT	3	MEUSE-ARGONNE (CHAMPAGNE)
371	4	SOMMEPY WOOD	1918	OCT	3	MEUSE-ARGONNE (CHAMPAGNE)
372	4	BLANC MONT II	1918	OCT	8	MEUSE-ARGONNE (CHAMPAGNE)
373	4	MEUSE-ARGONNE II	1918	OCT	4	MEUSE-ARGONNE PHASE II
374	4	EXERMONT-MONTREFAGNE	1918	OCT	4	MEUSE-ARGONNE PHASE II
375	4	MAYACHE RAVINE	1918	OCT	4	MEUSE-ARGONNE PHASE II
376	4	LA NEUVILLE LE COMTE FERME	1918	OCT	4	MEUSE-ARGONNE PHASE II
377	4	FERME DES GRANGES-FLEVILLE	1918	OCT	4	MEUSE-ARGONNE PHASE II
378	4	HILL 212	1918	OCT	5	MEUSE-ARGONNE PHASE II
379	4	BOIS DE BOYON-MONTREFAGNE	1918	OCT	5	MEUSE-ARGONNE PHASE II
380	4	HILL 272	1918	OCT	9	MEUSE-ARGONNE PHASE II
381	4	MEUSE-ARGONNE III	1918	NOV	1	MEUSE-ARGONNE PHASE III
382	4	REMILLY-AILLICOURT	1918	NOV	6	MEUSE-ARGONNE PHASE III
383	4	HILL 252-PONT MAUGIS	1918	NOV	7	MEUSE-ARGONNE PHASE III
384	4	THE PIAVE	1918	JUN	15	ITALIAN FRONT 1918
385	4	MEGIDDO	1918	SEP	19	PALESTINE 1918

ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
386	5	ALAM HALFA	1942	AUG	31	NORTH AFRICA 1942
387	5	EL ALAMEIN II	1942	OCT	23	NORTH AFRICA 1942
388	5	OPERATION LIGHTFOOT	1942	OCT	23	NORTH AFRICA 1942
389	5	ALAMEIN BRIDGEHEAD EXPANSION	1942	OCT	26	NORTH AFRICA 1942
390	5	OPERATION SUPERCHARGE	1942	NOV	2	NORTH AFRICA 1942
391	5	CHOUIGUI PASS	1942	NOV	26	TUNISIA 1942
392	5	EL GUETTAR	1943	MAR	23	TUNISIA 1943
393	5	SEDJANNE-BIZERTE	1943	APR	23	TUNISIA 1943
394	5	AMPHITHEATER	1943	SEP	9	SALERNO
395	5	PORT OF SALERNO	1943	SEP	9	SALERNO
396	5	SELE-CALORE CORRIDOR	1943	SEP	11	SALERNO
397	5	BATTIPAGLIA I	1943	SEP	12	SALERNO
398	5	VIETRI	1943	SEP	12	SALERNO
399	5	TOBACCO FACTORY	1943	SEP	13	SALERNO
400	5	BATTIPAGLIA II	1943	SEP	17	SALERNO
401	5	EBOLI	1943	SEP	17	SALERNO
402	5	VIETRI II	1943	SEP	17	SALERNO
403	5	GRAZZANISE	1943	OCT	12	VOLTURNO
404	5	CAIAZZO	1943	OCT	13	VOLTURNO
405	5	CAPUA	1943	OCT	13	VOLTURNO
406	5	CASTEL VOLTURNO	1943	OCT	13	VOLTURNO
407	5	MONTE ACERO	1943	OCT	13	VOLTURNO
408	5	TRIFLISCO	1943	OCT	13	VOLTURNO
409	5	DRAGONI	1943	OCT	15	VOLTURNO
410	5	CANAL I	1943	OCT	17	VOLTURNO
411	5	MONTE GRANDE (VOLTURNO)	1943	OCT	16	VOLTURNO
412	5	CANAL II	1943	OCT	18	VOLTURNO
413	5	FRANCOLISE	1943	OCT	20	VOLTURNO
414	5	SANTA MARIA OLIVETO	1943	NOV	4	VOLTURNO
415	5	MONTE CAMINO I	1943	NOV	5	VOLTURNO
416	5	MONTE LUNGO	1943	NOV	6	VOLTURNO
417	5	POZZILLI	1943	NOV	6	VOLTURNO
418	5	MONTE CAMINO II	1943	NOV	8	VOLTURNO
419	5	MONTE ROTONDO	1943	NOV	8	VOLTURNO
420	5	CALABRITTO	1943	DEC	1	VOLTURNO
421	5	MONTE CAMINO III	1943	DEC	2	VOLTURNO
422	5	MONTE MAGGIORE	1943	DEC	2	VOLTURNO
423	5	APRILIA I	1944	JAN	25	ANZIO
424	5	THE FACTORY	1944	JAN	27	ANZIO
425	5	CAMPOLEONE	1944	JAN	29	ANZIO
426	5	CAMPOLEONE COUNTERATTACK	1944	FEB	3	ANZIO
427	5	CARROCETO	1944	FEB	7	ANZIO

428	5	MOLETTA RIVER DEFENSE	1944	FEB	7	ANZIO
429	5	APPILIA II	1944	FEB	9	ANZIO
430	5	FACTORY COUNTERATTACK	1944	FEB	11	ANZIO
431	5	BOWLING ALLEY	1944	FEB	16	ANZIO
432	5	MOLETTA RIVER II	1944	FEB	16	ANZIO
433	5	FIOCCIA	1944	FEB	21	ANZIO
434	5	SANTA MARIA INFANTE	1944	MAY	12	ROME
435	5	SAN MARTINO	1944	MAY	12	ROME
436	5	CASTELLONORATO	1944	MAY	14	ROME
437	5	SPIGNO	1944	MAY	14	ROME
438	5	FORMIA	1944	MAY	16	ROME
439	5	MONTE GRANDE (ROME)	1944	MAY	17	ROME
440	5	ITRI-FONDI	1944	MAY	20	ROME
441	5	TERRACINA	1944	MAY	22	ROME
442	5	MOLETTA OFFENSIVE	1944	MAY	23	ROME
443	5	ANZIO-ALBANO ROAD	1944	MAY	23	ROME
444	5	ANZIO BREAKOUT	1944	MAY	23	ROME
445	5	CISTERNA	1944	MAY	23	ROME
446	5	SEZZE	1944	MAY	25	ROME
447	5	VELLETRI	1944	MAY	26	ROME
448	5	CAMPOLEONE	1944	MAY	26	ROME
449	5	VILLA CROCETTA	1944	MAY	27	ROME
450	5	ARDEA	1944	MAY	28	ROME
451	5	FOSSO DI CAMPOLEONE	1944	MAY	29	ROME
452	5	LANUVIO	1944	MAY	29	ROME
453	5	LARIANO	1944	JUN	1	ROME
454	5	VIA ANZIATE	1944	JUN	1	ROME
455	5	VALMONTONE	1944	JUN	1	ROME
456	5	TARTO-TIBER	1944	JUN	3	ROME
457	5	IL GIOGIO PASS	1944	SEP	13	NORTH ITALIAN
458	5	ST. LO	1944	JUL	11	NORMANDY
459	5	OPERATION GOODWOOD	1944	JUL	18	NORMANDY
460	5	OPERATION COBRA	1944	JUL	24	NORMANDY
461	5	MORTAIN	1944	AUG	6	NORMANDY BREAKOUT
462	5	CHARTRES	1944	AUG	16	LE MANS TO METZ
463	5	MELUN	1944	AUG	23	LE MANS TO METZ
464	5	SEINE RIVER	1944	AUG	23	LE MANS TO METZ
465	5	MOSELLE-METZ	1944	SEP	6	LE MANS TO METZ
466	5	METZ	1944	SEP	13	LE MANS TO METZ
467	5	ARRACOURT	1944	SEP	19	NORTHWEST EUROPE 1944
468	5	WESTWALL	1944	OCT	2	AACHEN
469	5	SCHMIDT	1944	NOV	2	NORTHWEST EUROPE 1944
470	5	SEILLE-NIED	1944	NOV	8	SAAR (LORRAINE)
471	5	FORET DE CHATEAU-SALINS	1944	NOV	10	SAAR (LORRAINE)
472	5	MORHANGE	1944	NOV	13	SAAR (LORRAINE)
473	5	MORHANGE-FAULQUEMONT	1944	NOV	13	SAAR (LORRAINE)
474	5	BOURGALTROFF	1944	NOV	14	SAAR (LORRAINE)
475	5	SARRE-ST. AVOLD	1944	NOV	20	SAAR (LORRAINE)
476	5	BAERENDORF I	1944	NOV	24	SAAR (LORRAINE)
477	5	BAERENDORF II	1944	NOV	26	SAAR (LORRAINE)
478	5	BURBACH-DURSTEL	1944	NOV	27	SAAR (LORRAINE)
479	5	DURSTEL-FAERBERSVILLER	1944	NOV	28	SAAR (LORRAINE)
480	5	SARRE-UNION	1944	DEC	1	SAAR (LORRAINE)
481	5	SARRE-SINGLING	1944	DEC	6	SAAR (LORRAINE)
482	5	SINGLING-BINING	1944	DEC	6	SAAR (LORRAINE)
483	5	SAUER RIVER	1944	DEC	16	ARDENNES
484	5	ST. VITH	1944	DEC	17	ARDENNES
485	5	BASTOGNE	1944	DEC	18	ARDENNES
ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
486	6	SEDAN-MEUSE RIVER	1940	MAY	13	FRANCE 1940
487	6	JITRA	1941	DEC	12	MALAYA 1941
488	6	ROVNO	1941	JUN	22	BARBAROSSA
489	6	DEFENSE OF MOSCOW	1941	SEP	30	TYPHOON
490	6	MOSCOW COUNTEROFFENSIVE	1941	DEC	5	MOSCOW COUNTEROFFENSIVE
491	6	POGORELOYE GORODISHCHE	1942	AUG	4	THE RZEHW OPERATION
492	6	LENINGRAD	1943	JAN	12	LENINGRAD
493	6	OBOYAN-KURSK (PHASE I)	1943	JUL	4	KURSK (CITADEL)
494	6	OPERATION CITADEL (SOUTHERN SECTOR)	1943	JUL	5	KURSK (CITADEL)
495	6	OBOYAN-KURSK (PHASE II)	1943	JUL	7	KURSK (CITADEL)
496	6	OBOYAN-KURSK (PHASE III)	1943	JUL	11	KURSK (CITADEL)
497	6	PROKHOROVKA	1943	JUL	12	KURSK (CITADEL)
498	6	KURSK COUNTEROFFENSIVE	1943	AUG	3	KURSK COUNTEROFFENSIVE

499	6	BELGOROD	1943 AUG 3	KURSK COUNTEROFFENSIVE
500	6	MELITOPOL	1943 SEP 26	RACE TO THE DNIEPER
501	6	KORSUN-SCHEVCHENKOVSKIY	1944 JAN 24	UKPANIAN CAMPAIGN
502	6	NIKOPOL BRIDGEHEAD	1944 JAN 31	DNIEPER BRIDGEHEAD
503	6	SEVASTOPOL	1944 MAY 5	CRIMEA 1944
504	6	BEPEZINA RIVER	1944 JUN 25	BYFLORUSSIAN OFFENSIVE
505	6	LVOV-SANDOMIERZ	1944 JUL 13	LIBERATION OF E POLAND
506	6	BRODY (PHASE I)	1944 JUL 14	THE LVOV-SANDOMIERZ OPERATION
507	6	BRODY (PHASE II)	1944 JUL 15	THE LVOV-SANDOMIERZ OPERATION
508	6	VISTULA RIVER (PHASE D I)	1944 JUL 29	POLAND 1944
509	6	VISTULA RIVER (PHASE D II)	1944 AUG 2	POLAND 1944
510	6	YASSY-KISHINEV	1944 AUG 20	ROMANIA 1944
511	6	VISTULA-ODER	1945 JAN 12	LIBERATION OF W POLAND
512	6	EAST PRUSSIA	1945 JAN 13	EAST PRUSSIA 1945
513	6	CIECHANOW (PHASE I)	1945 JAN 14	RUSSIAN WINTER OFFENSIVE 1945
514	6	CIFCHANOW (PHASE II)	1945 JAN 15	RUSSIAN WINTER OFFENSIVE 1945
515	6	SEFLOW HEIGHTS	1945 APR 16	BERLIN 1945
516	6	MUTANKIANG	1945 AUG 9	MANCHURIA 1945
517	6	TARAWA-BETIO	1944 NOV 20	CENTRAL PACIFIC
518	6	IWO JIMA - INTO THE MAIN DEFENSES	1945 FEB 20	BONIN ISLANDS
519	6	IWO JIMA - SURIBACHI	1945 FEB 20	BONIN ISLANDS
520	6	IWO JIMA - FINAL PHASE	1945 MAR 11	BONIN ISLANDS
521	6	ADVANCE FROM THE BEACH	1945 APR 2	OKINAWA (US 7TH INF DIV SECTOR)
522	6	ADVANCE THROUGH THE OUTPOSTS	1945 APR 5	OKINAWA (US 7TH INF DIV SECTOR)
523	6	TOMB HILL-OUKI	1945 APR 9	OKINAWA (US 7TH INF DIV SECTOR)
524	6	SKYLINE RIDGE-ROCKY CRAGS	1945 APR 19	OKINAWA (US 7TH INF DIV SECTOR)
525	6	KOCHI RIDGE-ONAGA I	1945 APR 25	OKINAWA (US 7TH INF DIV SECTOR)
526	6	KOCHI RIDGE-ONAGA II	1945 APR 28	OKINAWA (US 7TH INF DIV SECTOR)
527	6	KOCHI RIDGE-ONAGA III	1945 APR 30	OKINAWA (US 7TH INF DIV SECTOR)
528	6	JAPANESE COUNTERATTACK	1945 MAY 4	OKINAWA (US 7TH INF DIV SECTOR)
529	6	KOCHI RIDGE IV	1945 MAY 6	OKINAWA (US 7TH INF DIV SECTOR)
530	6	SHURI ENVELOPMENT (PHASE I)	1945 MAY 22	OKINAWA (US 7TH INF DIV SECTOR)
531	6	JAPANESE COUNTERATTACKS	1945 MAY 24	OKINAWA (US 7TH INF DIV SECTOR)
532	6	SHURI ENVELOPMENT (PHASE II)	1945 MAY 26	OKINAWA (US 7TH INF DIV SECTOR)
533	6	SHURI ENVELOPMENT (PHASE III)	1945 MAY 29	OKINAWA (US 7TH INF DIV SECTOR)
534	6	HILL 95-I	1945 JUN 6	OKINAWA (US 7TH INF DIV SECTOR)
535	6	HILL 95-II	1945 JUN 9	OKINAWA (US 7TH INF DIV SECTOR)
536	6	YAFJU-DAKE	1945 JUN 12	OKINAWA (US 7TH INF DIV SECTOR)
537	6	HILLS 153 AND 115	1945 JUN 15	OKINAWA (US 7TH INF DIV SECTOR)
538	6	ADVANCE FROM THE BEACH	1945 APR 2	OKINAWA (US 96TH INF DIV SECTOR)
539	6	ADVANCE TO SHURI LINE OUTPOSTS	1945 APR 5	OKINAWA (US 96TH INF DIV SECTOR)
540	6	KAKAZU AND TOMBSTONE RIDGES	1945 APR 9	OKINAWA (US 96TH INF DIV SECTOR)
541	6	NISHIBARU RIDGE-TANABARU ESCARPMENT	1945 APR 19	OKINAWA (US 96TH INF DIV SECTOR)
542	6	MAFDA ESCARPMENT	1945 APR 26	OKINAWA (US 96TH INF DIV SECTOR)
543	6	ATTACK ON SHURI LINE'S EASTERN FLANK I	1945 MAY 11	OKINAWA (US 96TH INF DIV SECTOR)
544	6	ATTACK ON SHURI LINE'S EASTERN FLANK II	1945 MAY 14	OKINAWA (US 96TH INF DIV SECTOR)
545	6	ATTACK ON SHURI LINE'S EASTERN FLANK III	1945 MAY 20	OKINAWA (US 96TH INF DIV SECTOR)
546	6	ADVANCE TO YUZA-DAKE/YAEJU-DAKE ESCARPMENT	1945 JUN 6	OKINAWA (US 96TH INF DIV SECTOR)
547	6	ATTACK ON YUZA-DAKE/YAEJU-DAKE ESCARPMENT	1945 JUN 10	OKINAWA (US 96TH INF DIV SECTOR)
548	6	CAPTURE OF YUZA-DAKE/YAEJU-DAKE ESCARPMENT	1945 JUN 12	OKINAWA (US 96TH INF DIV SECTOR)
549	6	JENIN	1967 JUN 5	WEST BANK
550	6	JERUSALEM	1967 JUN 5	WEST BANK
551	6	KABATIYA	1967 JUN 6	WEST BANK
552	6	TILFIT-ZABABIBA	1967 JUN 6	WEST BANK
553	6	NARLUS	1967 JUN 7	WEST BANK
554	6	RAFAH	1967 JUN 5	SINAI
555	6	BIR LAHFAN	1967 JUN 5	SINAI
556	6	ABU AGEILA-UM KATEF	1967 JUN 5	SINAI
557	6	EL-ARISH	1967 JUN 5	SINAI
558	6	JEBEL LIBNI	1967 JUN 6	SINAI
559	6	GAZA STRIP	1967 JUN 6	SINAI
560	6	BIP HASSNA-BIR THAMADA	1967 JUN 7	SINAI
561	6	MITLA PASS	1967 JUN 7	SINAI
562	6	BIR HAMA-BIR GIFGAFA	1967 JUN 7	SINAI
563	6	NAKHL	1967 JUN 8	SINAI
564	6	BIP GIFGAFA	1967 JUN 8	SINAI
565	6	TEL FAHAR-BANIAS	1967 JUN 9	GOLAN
566	6	RAWIYEH	1967 JUN 9	GOLAN
567	6	ZAOURA-KALA	1967 JUN 9	GOLAN
568	6	KERAMA	1968 MAR 21	JORDAN VALLEY
569	6	SUFZ CANAL ASSAULT, NORTH	1973 OCT 6	SUFZ 1973

ISEQNO	VOLNO	NAME	YEAR	MON	DA	CAMPGN
570	6	SUEZ CANAL ASSAULT, SOUTH	1973	OCT	6	SUEZ 1973
571	6	SECOND ARMY BUILDUP	1973	OCT	7	SUEZ 1973
572	6	THIRD ARMY BUILDUP	1973	OCT	7	SUEZ 1973
573	6	KANTARA-FIRDAN	1973	OCT	8	SUEZ 1973
574	6	EGYPTIAN OFFENSIVE, NORTH	1973	OCT	14	SUEZ 1973
575	6	EGYPTIAN OFFENSIVE, SOUTH	1973	OCT	14	SUEZ 1973
576	6	DEVERSOIR (CHINESE FARM I)	1973	OCT	15	SUEZ 1973
577	6	DEVERSOIR (CHINESE FARM II)	1973	OCT	16	SUEZ 1973
578	6	DEVERSOIR WEST	1973	OCT	18	SUEZ 1973
579	6	ISMAILIA	1973	OCT	19	SUEZ 1973
580	6	JEREL GENEIFA	1973	OCT	19	SUEZ 1973
581	6	SHALLUFA I	1973	OCT	22	SUEZ 1973
582	6	SHALLUFA II	1973	OCT	23	SUEZ 1973
583	6	SUEZ	1973	OCT	23	SUEZ 1973
584	6	ADABIYA	1973	OCT	23	SUEZ 1973
585	6	KUNEITRA	1973	OCT	6	GOLAN
586	6	AHMADIYEH	1973	OCT	6	GOLAN
587	6	RAFID	1973	OCT	6	GOLAN
588	6	YEHUDA-EL AL	1973	OCT	7	GOLAN
589	6	NAFEKH	1973	OCT	7	GOLAN
590	6	TEL FARRIS	1973	OCT	8	GOLAN
591	6	HUSHNIYAH	1973	OCT	8	GOLAN
592	6	MOUNT HERMONIT	1973	OCT	8	GOLAN
593	6	MOUNT HERMON I	1973	OCT	8	GOLAN
594	6	TEL SHAMS	1973	OCT	11	GOLAN
595	6	TEL SHAAR	1973	OCT	11	GOLAN
596	6	TEL EL HARA	1973	OCT	13	GOLAN
597	6	KFAR SHAMS-TEL ANTAR	1973	OCT	15	GOLAN
598	6	NARA	1973	OCT	16	GOLAN
599	6	ARAB COUNTEROFFENSIVE	1973	OCT	19	GOLAN
600	6	MOUNT HERMON II	1973	OCT	21	GOLAN
601	6	MOUNT HERMON III	1973	OCT	22	GOLAN

APPENDIX I

DESCRIPTION OF CDES CONTRACT TASKS

I-1. OBJECTIVE. The purpose of the CDES contract is to correct typographical mistakes, omissions, inconsistencies, and ambiguities in the battle and engagement data base being used in the CHASE Study.

I-2. BACKGROUND

a. In 1983 and 1984, the Historical Evaluation and Research Organization (HERO), under contract MDA903-82-C-0363, prepared for the US Army Concepts Analysis Agency (CAA) a detailed data base of battles and engagements. In September 1984, CAA published this as "Analysis of Factors That Have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements," Study Report CAA-SR-84-6.

b. In accordance with the previous contract, the data base was detailed for individual battles. It is not, however, directly usable in Army studies and analyses, tactical concept formulation, or wargaming. These activities require summary, quantitative relationships applicable throughout a broad range of engagement situations to identify significant trends or factors. In August 1984, CAA initiated the Combat History Analysis Study Effort (CHASE) to search the HERO data base for historically based quantitative relationships for use in Army studies and analyses, concept formulation, and wargaming. The CHASE Study has identified a need for extending the original research effort to make the data base useful for other analyses.

I-3. SCOPE OF THE CDES CONTRACT

a. General. The tasks to be addressed by the contractor are described in paragraphs b through j below. In addition, a final report in the form of an errata addendum is required. The addendum package should document the results of the tasks listed below. The package should also be distributable to current holders of the original data base.

b. Task 1. Analyze Data Base Problem Reports.

(1) Background. CAA has compiled a list of problem reports as it transcribed the HERO data base into computerized format for use in CHASE. The problem reports identify typographical mistakes, missing, ambiguous, or suspect data items and terminology in the HERO data base. Problem reports have been accumulated, and their resolution is required to use the data base effectively.

(2) Task Statement. Review each of the problem reports and correct or clarify the data base as required to resolve the problem identified.

c. **Task 2.** Clarify the Total Engaged Personnel Strength Data.

(1) **Background.** HERO defines (see Appendix E) the total engaged personnel strength to be "The sum, at the start of the engagement, of all personnel subject to enemy fire, including generally combat and combat support troops but also service troops if subject to enemy fire." However, "For lengthy engagements in which both sides were significantly reinforced after the beginning of the engagement, an average of the daily start strength(s) was entered." The differences in these definitions of total engaged personnel strength explain why, in some instances, the casualties can exceed the "total engaged" personnel strength. Neither the initial strengths, nor the total reinforcements/replacements, nor the final personnel strengths can be recovered from the data provided by HERO. Also, the data base does not indicate whether the total engaged are the initial strengths or daily averages.

(2) **Task Statement.** Identify the battles for which the total engaged strength represents the number of personnel at the start of the battle or an average daily strength during the battle. Explain the derivation of each average daily strength computation and provide the initial personnel strengths, the number of reinforcements/replacements, and the final personnel strengths for those battles.

d. **Task 3.** Clarify the Basis for Assigning Victory.

(1) **Background.** Hero states (see Appendix E) that "the victor, if not apparent from the decisive resolution of the combat in favor of one side or the other, is determined by an assessment of the extent to which each side was successful in accomplishing its mission." Thus, two distinct criteria for assigning victory were used, but the data base does not indicate which criterion applies to the victory.

(2) **Task Statement.** Identify the battles for which the victory was assigned on the basis of "the decisive resolution of combat in favor of one side or the other," and those for which it was assigned on the basis of "the extent to which each side was successful in accomplishing its mission."

e. **Task 4.** Refine the Duration Data.

(1) **Background.** The HERO data base lists battle duration in days, but this time scale is too coarse to be readily usable for CAA studies and analyses. Battles that last for less than a day, or span just 2 or 3 days, have durations that are badly misrepresented by this coarse a time scale. For example, suppose two battles occur with identical personnel strengths, casualty losses, and distance advance; but that the first battle lasts 1.5 hours while the second battle lasts 15 hours. Both would be listed in the HERO data base as having the same percent casualties per day and the same rate of advance per day. Yet the first battle actually had casualty and advance rates 10 times those of the second battle.

(2) Task Statement. Identify the battles for which a refined and more accurate value of battle duration can be assigned, and restate the duration of those battles. For example, if the time data available for a particular battle indicates "the battle lasted from sunrise to sunset during August," then modify the battle duration and indicate the new time in the addendum.

f. Task 5. Clarify the Width of Front Data.

(1) Background. HERO states (see Appendix E) with regard to the width of a front that "where there is a significant difference between the fronts occupied by the opposing forces in an engagement, the width of the attacker's front is entered as the descriptor." However, the data base does not indicate when the width of front applies to the defender as well as to the attacker.

(2) Task Statement. Provide the defender's width of front for all battles.

g. Task 6. Clarify the Defender Posture Description.

(1) Background. HERO states (see Appendix E) with regard to the defender posture data that "frequently, it should be noted, descriptors entered in this category reflect a defensive posture best defined as a combination or average of 2 of the 5 basic categories. For example, a defender may adopt two postures during the course of an engagement, or the level of defensive preparation may not be uniform across a lengthy front or throughout the depth of a defended zone." However, the data base does not indicate when the descriptors identify a combination or average of the basic categories.

(2) Task Statement. Identify those battles for which the defender posture indicates a "combination" descriptor, and those for which it indicates an "average" descriptor. Also, state whether the changes in defensive postures which warranted the modified descriptor occurred along the front, depth, or time of the defense. For example, if an average descriptor is listed due to significant changes along the defensive front, indicate that fact adjacent to the modified descriptor.

h. Task 7. Identify the Quality of Strength and Loss Data.

(1) Background. Some of the data within the data base are more reliable than others; however, the HERO data base does not indicate the level of confidence that can be assigned to the data. Assigning a "weight" indicating the adjudged relative level of reliability of the data would be very useful for certain statistical analysis purposes. It is probably inappropriate to assign relative reliability weights to values such as those in Tables 2, 4, 5, and 6 that are themselves judgmental in nature. However, it would be appropriate to assign relative reliability weights to objective values such as those in Table 3 which contain strength and loss data.

(2) Task Statement. Assign to each battle a "weight" that indicates the adjudged relative level of reliability for the strength and loss data of the respective battles.

i. Task 8. Develop Strength and Attrition Histories for Selected Battles.

(1) Background. The HERO data base provides data on the total engaged strengths and losses experienced in each historical battle. While useful for many purposes, these data cannot be used to study the laws governing attrition in historical battles. What is required are data listing the personnel strength and cumulative attrition at intermediate times during the course of a battle. This type of data were used by Engel and by Busse in their classical analyses. Augmentation of the HERO data base to provide attrition histories for selected battles would allow a considerably deeper analysis of attrition to be performed than is possible without it.

(2) Task Statement. List those battles where accurate strength and attrition histories are available for both sides. Select a list of battles based upon CAA approval for which two-sided strengths and attrition histories will be prepared. Develop and document the strength and attrition histories for each of these battles.

j. Task 9. Assistance in Eliminating Unwanted Redundancies.

(1) Background. Tables 2, 4, 5, 6, and 7 of the HERO data base contain at least 28 columns of information, some of which seems to be redundant. For example, Table 2 gives information on "whether or not surprise (was) achieved by one side or the other; and if it had been, by whom and to what degree." Table 4 contains columns characterizing the disparity between the opponents with respect to such items as leadership, combat effectiveness, and military intelligence. Table 6 categorizes the extent to which the battle outcome was affected by such factors as leadership, planning, surprise, and maneuver. There seems to be a high degree of redundancy among all of the factors mentioned above. For technical statistical reasons, it is necessary to reduce these data to a much smaller number of columns that capture the gist of the information without redundant information. CAA expects to use statistical methods to assist in this reduction process; however, historical insights may have a valuable role to play in this process.

(2) Task Statement. Review CAA's efforts to reduce the level of redundancy in the data based upon the data judgements integrated into those tables. Identify points of concern, and suggest appropriate methods for accomplishing the removal of redundancy without losing essential information.

APPENDIX J

AN INTRODUCTION TO LOGISTIC REGRESSION

J-1. INTRODUCTION. Suppose that we have N observations as follows:

$$
\begin{array}{llllll}
y_1, & x_{10}, & x_{11}, & x_{12}, & \cdots, & x_{1P} \\
y_2, & x_{20}, & x_{21}, & x_{12}, & \cdots, & x_{2P} \\
y_3, & x_{30}, & x_{31}, & x_{32}, & \cdots, & x_{3P} \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
y_N, & x_{N0}, & x_{N1}, & x_{N2}, & \cdots, & x_{NP}
\end{array}
$$

where each of the y_n for n = 1(1)N is one* of the integer values in the set of response levels r = 0(1)R and x_{np} is a real number for n = 1(1)N and p = 0(1)P. We assume that each y_n is the result of an experimental trial in which the value of y_n is selected randomly from the values 0(1)R according to the probabilities;

$$\text{Prob}(y_n = r) = P_r(\underline{x}_n) \tag{J-1}$$

where

$$\underline{x}_n = (x_{n0}, x_{n1}, \cdots, x_{nP})$$

is a (P+1)-dimensional array of real numbers that characterizes the conditions under which the n-th experimental trial was conducted.

a. Example. When all of the experimental trials are conducted under identical conditions, then all of the $\underline{x}_n$'s are equal, i.e., for n = 1(1)N we have $\underline{x}_n = \underline{x}_0$. Then we also have

$$P_r(\underline{x}_n) = P_r(\underline{x}_0) = \text{(say) } P_r$$

---

*In this paper the notation u(v)w is used to stand for the set of values u, u+v, u+2v, ... , w.

for all n = 1(1)N. Hence, this case reduces to the well-known multinomial situation, i.e., if we let N_r be the number of times response level r occurred, then the N_r will in this special case be distributed according to a joint multinomial distribution, i.e.,

$$\text{Prob}(N_r = n_r \text{ for } r = 0(1)R) = \prod_{r=0}^{R} \frac{P_r^{n_r}}{n_r!}$$

where the n_r are constrained by the identity

$$\sum_{r=0}^{R} n_r = N,$$

where N is the total number of observations.

b. In many applications it is appropriate to take $x_{n0} = 1$ for n = 1(1)N. This corresponds to allowing a nonzero "intercept" in ordinary linear regression.

J-2. SPECIALIZATION TO THE LOGISTIC CASE. Various results follow from different assumed functional forms for the $P_r(\underline{x}_n)$. In this appendix, we shall use only the logistic form:

$$P_r(\underline{x}_n) = \frac{N_r^*(\underline{x}_n)}{D^*(\underline{x}_n)} \tag{J-2}$$

where

$$D^*(\underline{x}_n) = \sum_{r=0}^{R} N_r^*(\underline{x}_n), \tag{J-3}$$

$$N_r^*(\underline{x}_n) = \exp(\underline{a}_r^* \cdot \underline{x}_n) = \exp\left(\sum_{p=0}^{P} a_{rp}^* x_{np}\right), \tag{J-4}$$

and each $\underline{a}_r^*$ is a (P+1)-dimensional array of real-valued parameters whose values are unknown, and which may be chosen to fit closely the experimental observations $(\underline{y},X)$, where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix}$$

and

$$X = \begin{bmatrix} x_{10} & x_{11} \cdots x_{1P} \\ x_{20} & x_{21} \cdots x_{2P} \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ x_{N0} & x_{N1} \cdots x_{NP} \end{bmatrix} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_N \end{bmatrix} .$$

Note that by Equation (J-4) the numerator $N_0^*(\underline{x}_n)$ is an exponential, and hence it is always greater than zero, so we can divide both numerator and denominator of the expression for $P_r(\underline{x}_n)$ by $N_0^*(\underline{x}_n)$ in Equation (J-2) and thus write for r = 0(1)R:

$$P_r(\underline{x}_n) = \frac{N_r(\underline{x}_n)}{D(\underline{x}_n)} \tag{J-5}$$

where

$$D(\underline{x}_n) = D^*(\underline{x}_n)/N_0^*(\underline{x}_n) = 1 + \sum_{r=1}^{R} N_r(\underline{x}_n), \tag{J-6}$$

$$N_r(\underline{x}_n) = N_r^*(\underline{x}_n)/N_0^*(\underline{x}_n) = \exp\ (\underline{a}_r \cdot \underline{x}_n) = \exp\ \left(\sum_{p=0}^{P} a_{rp}x_{np} \right), \tag{J-7}$$

and each

$$\underline{a}_r = \underline{a}_r^* - \underline{a}_0^* \tag{J-8}$$

for r = 0(1)R is a (P+1)-dimensional array of real numbers with the special feature that $\underline{a}_0 = \underline{0}$.

Because the $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_R$ completely determine the $P_r(\underline{x}_n)$ values, they will be called the essential parameters, and the response levels r = 1(1)R will be called the essential response levels.

J-3. LIKELIHOOD FUNCTION. We will now establish the likelihood function for this situation.

a. To do that we first define for r = 0(1)R and n = 1(1)N the indicator function

$$\begin{aligned} e_{rn} &= 1, \text{ if } y_n = r, \text{ and} \\ e_{rn} &= 0, \text{ otherwise.} \end{aligned} \tag{J-9}$$

Evidently the indicator function e_{rn} has the following properties.

(1) For r = 0(1)R, $\sum_{n=1}^{N} e_{rn} = N_r$, (J-10)

where N_r is the number of y_n's that are equal to r.

(2) For n = 1(1)N, $\sum_{r=0}^{R} e_{rn} = 1$, (J-11)

because each y_n must be one, and only one, of the values 0(1)R.

(3) $\sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} = \sum_{r=0}^{R} N_r = N$. (J-12)

b. Using the indicator function we can express the logarithm of the likelihood as a function of the essential parameters. Specifically, the log-likelihood function will be:

$$L(\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_R) = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \,\text{LOG}(P_r(\underline{x}_n)). \tag{J-13}$$

When $\underline{a}_r = \underline{0}$ for r = 1(1)R, we have $N_r(\underline{x}_n) = 1$ for r = 1(1)R, and $D(\underline{x}_n) = 1+R$, so that $P_r(\underline{x}_n) = (1+R)^{-1}$ for r = 0(1)R.

Accordingly we have:

$$L(\underline{0}, \underline{0}, \ldots , \underline{0}) = - \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \, LOG(1+R) = -N * LOG(1+R). \qquad (J\text{-}14)$$

J-4. INFORMATION MATRIX

a. The derivative of L with respect to the parameter a_{st} will now be determined, where s = 1(1)R and t = 0(1)P. To do this we proceed as follows:

$$L(\underline{a}_1, \underline{a}_2, \ldots , \underline{a}_R) = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \, LOG(P_r(\underline{x}_n))$$

$$= \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \, (LOG(N_r(\underline{x}_n)) - LOG(D(\underline{x}_n))). \qquad (J\text{-}15)$$

so for s = 1(1)R and t = 0(1)P we have:

$$\frac{dL}{da_{st}} = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \left[\frac{1}{N_r(\underline{x}_n)} \frac{dN_r(\underline{x}_n)}{da_{st}} - \frac{1}{D(\underline{x}_n)} \frac{dD(\underline{x}_n)}{da_{st}} \right].$$

But

$$\frac{dN_r(\underline{x}_n)}{da_{st}} = \frac{d}{da_{st}} \left[\exp \left(\sum_{p=0}^{P} a_{rp} x_{np} \right) \right]$$

$$= N_r(\underline{x}_n) \sum_{p=0}^{P} x_{np} \frac{da_{rp}}{da_{st}}$$

$$= g_{rs} \, N_r(\underline{x}_n) \, x_{nt}, \qquad (J\text{-}16)$$

where g_{rs} is Kronecker's delta-function, defined by $g_{rs} = 1$ if $r = s$, and $g_{rs} = 0$ otherwise. Also

$$\frac{dD(\underline{x}_n)}{da_{st}} = \frac{d}{da_{st}} \left[1 + \sum_{r=1}^{R} N_r(\underline{x}_n) \right]$$

$$= \sum_{r=1}^{R} \frac{dN_r(\underline{x}_n)}{da_{st}} = \sum_{r=1}^{R} g_{rs}\, N_r(\underline{x}_n)\, x_{nt}$$

$$= N_s(\underline{x}_n)\, x_{nt}. \qquad \text{(J-17)}$$

Thus:

$$\frac{dL}{da_{st}} = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \left[g_{rs} - \frac{N_s(\underline{x}_n)}{D(\underline{x}_n)} \right] x_{nt}$$

$$= \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \left[g_{rs} - P_s(\underline{x}_n) \right] x_{nt}$$

$$= \sum_{n=1}^{N} \left[\sum_{r=0}^{R} g_{rs}\, e_{rn} - P_s(\underline{x}_n) \sum_{r=0}^{R} e_{rn} \quad x_{nt} \right]$$

$$= \sum_{n=1}^{N} (e_{sn} - P_s(\underline{x}_n))\, x_{nt}\ , \quad \text{for } s = 1(1)R \text{ and } t = 0(1)P. \qquad \text{(J-18)}$$

We observe that when $\underline{a}_r = \underline{0}$ for $r = 1(1)R$, then

$P_s(\underline{x}_n) = (1+R)^{-1}$, and hence

$$\frac{dL}{da_{st}}(\underline{0}, \underline{0}, \ldots, \underline{0}) = \sum_{n=1}^{N} \left[e_{pn} - \frac{1}{1+R} \quad x_{nt} \right] \tag{J-19}$$

b. The information matrix is defined to be:

$$-E\left[\frac{d^2L}{d\theta_u d\theta_v}\right] = (V_{uv})^{-1} \tag{J-20}$$

where the maximum likelihood parameters are θ_u and θ_v. The covariance matrix of the θ's is $V = (V_{uv})$, the inverse of the information matrix (see, for example, **Ref J-1**, Volume 2, 57; **Ref J-2**; **Ref J-3**, page 87).

In our case we already know from Equation (J-18) that

$$\frac{dL}{da_{st}} = \sum_{n=1}^{N} (e_{sn} - P_s(\underline{x}_n)) \quad x_{nt}, \text{ for } s = 1(1)R \text{ and } t = 0(1)P.$$

Changing s to r, and t to p, we write this as:

$$\frac{dL}{da_{rp}} = \sum_{n=1}^{N} (e_{rn} - P_r(\underline{x}_n)) \quad x_{np}.$$

Then

$$\frac{d^2L}{da_{st}da_{rp}} = - \sum_{n=1}^{N} x_{np} \frac{dP_r(\underline{x}_n)}{da_{st}}$$

$$= - \sum_{n=1}^{N} x_{np} \left[\frac{dN_r(\underline{x}_n)}{da_{st}} \frac{1}{D(\underline{x}_n)} - \frac{N_r(\underline{x}_n)}{D(\underline{x}_n)^2} \frac{dD(\underline{x}_n)}{da_{st}} \right] .$$

Substituting from Equation (J-16) and (J-17) yields the information matrix element $V_{(s,t)(r,p)}$ as:

$$\frac{d^2L}{da_{st}da_{rp}} = - \sum_{n=1}^{N} x_{np} \left[\frac{g_{rs}N_r(\underline{x}_n)x_{nt}}{D(\underline{x}_n)} - \frac{N_r(\underline{x}_n)}{D(\underline{x}_n)^2} N_s(\underline{x}_n)x_{nt} \right]$$

$$= - \sum_{n=1}^{N} x_{np}x_{nt}P_r(\underline{x}_n) \; (g_{rs} - P_s(\underline{x}_n)) . \qquad \text{(J-21)}$$

where s,r = 1(1)R and t,p = 0(1)P.

Note that we also have from Equations (J-14) and (J-18):

$$L(\underline{0}) = - N * \mathrm{LOG}(1+R), \text{ and} \qquad \text{(J-22)}$$

$$\frac{dL}{da_{rp}} = \sum_{n=1}^{N} (e_{rn} - P_r(x_n)) \; x_{np} \qquad \text{(J-23)}$$

for r = 1(1)R and p = 0(1)P.

Equations (J-21), (J-22), and (J-23) are the results sought.

c. Additional information on logistic and probit regression can be obtained from books by Cox, Aldrich and Daganzo **(Refs J-3, J-4, J-5).**

J-5. A NUMERICAL EXAMPLE. Suppose that we wish to fit a logistic function to the data given in Table J-1. For these hypothetical data we have N = 10 observations, R = 1 essential response level, and P + 1 = 2 explanatory variables per response level. However, since for all observations x_{n0} = 0.0, only one explanatory variable actually appears. That is, we look for a fit of the form:

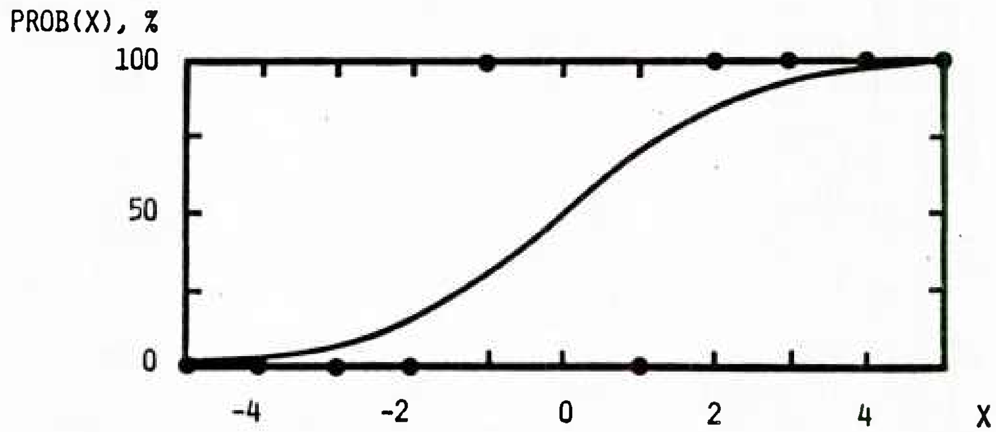
$$P_o(x) = \frac{1}{1 + \exp(a^*x)}$$

$$P_1(x) = 1 - P_o(x),$$

where $P_1(s)$ is the probability that the response will be 1 when the stimulus is x. Figure J-1 shows the hypothetical observations and the maximum likelihood of logistic fit. As indicated in Figure J-1, the maximum likelihood value of a is 0.862.

Table J-1. Hypothetical Data

Observation	y_n	x_{n0}	x_{n1}
1	1	0.0	-1.0
2	0	0.0	1.0
3	1	0.0	2.0
4	1	0.0	3.0
5	1	0.0	4.0
6	1	0.0	5.0
7	0	0.0	-2.0
8	0	0.0	-3.0
9	0	0.0	-4.0
10	0	0.0	-5.0



$$PROB(X) = \frac{EXP\ (0.862\ X)}{1 + EXP\ (0.862\ X)}$$

Figure J-1. Example of Logistic Regression Curve Fitted to 10 Hypothetical Data Points

J-6. SUMMARY OF RANGES OF VARIABLES

N = Number of sample points (i.e., the number of experimental trials)

P+1 = Number of explanatory variables per response level

P = Number of parameters per essential parameter-set

y_n = response to n-th stimulus, where n = 1(1)N

x_{np} = p-th component of the n-th stimulus, where n = 1(1)N and p = 0(1)P

x_{n0} = 1 for n = 1(1)N (usually)

R+1 = Number of possible values or response levels of the y_n's

R = Number of essential parameter sets, that is, the number of essential response levels

a_{rp} = The essential parameter associated with essential response level r and parameter p where r = 1(1)R and p = 0(1)P

$a_{op} = 0$ for $p = 0(1)P$ (by definition of $a_{rp} = a^*_{rp} - a^*_{op}$)

$e_{rn} = 1$ if $y_n = r$, and $e_{rn} = 0$ otherwise, where $r = 0(1)R$, $n = 1(1)N$

$$L = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn} \, \mathrm{LOG}(P_r(\underline{x}_n)) = \sum_{n=1}^{N} \sum_{r=0}^{R} e_{rn}(\mathrm{LOG}(N_r(\underline{x}_n)) - \mathrm{LOG}(D(\underline{x}_n)))$$

$$P_r(\underline{x}_n) = \frac{N_r(\underline{x}_n)}{D(\underline{x}_n)} \quad \text{for } r = 0(1)R \text{ and } n = 1(1)N$$

$$N_r(\underline{x}_n) = \exp\left(\sum_{p=0}^{P} a_{rp}x_{np}\right) \quad \text{for } r = 0(1)R \text{ and } n(1)N$$

$$D(\underline{x}_n) = \sum_{r=0}^{R} N_r(\underline{x}_n) \quad \text{for } n = 1(1)N$$

When $\underline{a}_1 = \underline{0}$, $\underline{a}_2 = \underline{0}$, ... , etc., then $P_r(\underline{x}_n) = (1+R)^{-1}$ for $r = 0(1)R$ and $n = 1(1)N$

$$L(\underline{0}) = -N \,\mathrm{LOG}\,(1+R)$$

$$\frac{dL}{da_{rp}} = \sum_{n=1}^{N} (e_{rn} - P_r(\underline{x}_n))\, x_{np}, \quad \text{for } r = 1(1)R \text{ and } p = 0(1)P.$$

APPENDIX K

DISTRIBUTION

Addressee	No of copies
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZA Washington, DC 20310	1
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZD Washington, DC 20310	1
Deputy Chief of Staff for Personnel Headquarters, Department of the Army ATTN: DAPE-ZA Washington, DC 20310	1
Deputy Chief of Staff for Logistics Headquarters, Department of the Army ATTN: DALO-ZA Washington, DC 20310	1
Commander US Army Logistics Center Fort Lee, VA 23801	1
Deputy Chief of Staff for Research, Development, and Acquisition Headquarters, Department of the Army ATTN: DAMA-ZA Washington, DC 20310	2
Deputy Under Secretary of the Army (Operations Research) Washington, DC 20310	4

Addressee	No of copies
Chief of Staff, Army ATTN: DACS-DMO Washington, DC 20310	1
Assistant Secretary of the Army (Research, Development, and Acquisition) Washington, DC 20310	1
Office of The Surgeon General ATTN: DASG-PSC 5111 Leesburg Pike Falls Church, VA 22041-3258	1
Commander National Guard Bureau Room 2E394 The Pentagon Washington, DC 20310	1
Director US Army TRADOC Analysis Center White Sands Missile Range, NM 88002	2
Director US Army Materiel Systems Analysis Agency Aberdeen Proving Ground, MD 21005	2
Director US Army Ballistic Research Laboratories Building 305 Aberdeen Proving Ground, MD 21005	2
Commander Combined Arms Combat Development Activity Fort Leavenworth, KS 66027	2

Addressee	No of copies
Commander Foreign Science and Technology Center 227th Street NE Charlottesville, VA 22901	1
Director Defense Nuclear Agency ATTN: LASS 6801 Telegraph Road Alexandria, VA 20305	2
Commander Army Research Institute 5001 Eisenhower Avenue Alexandria, VA 22333	3
Commander US Army Military Personnel Center 200 Stovall Street Alexandria, VA 22332	1
Commander US Army Troop Support Agency Fort Lee, VA 23801	1
Commander US Army Logistics Evaluation Agency New Cumberland Army Depot New Cumberland, PA 17070	1
Director Defense Logistics Studies Information Exchange US Army Logistics Management Center Fort Lee, VA 23801	1
Defense Technical Information Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22314-6145	2

Addressee	No of copies
US Army Nuclear and Chemical Agency 7500 Backlick Road, Bldg #2073 ATTN: Library Springfield, VA 22150	1
Commander Air Mobility Research and Development Laboratory Eustis Directorate Fort Eustis, VA 23604	1
Commander US Army Mobility Equipment Research and Development Command Fort Belvoir, VA 22060	1
Commander US Army Research, Development, and Acquisition Information Systems Agency Radford, VA 24141	1
The Pentagon Library (Army Studies Section) ATTN: ANRAL-RS The Pentagon Washington, DC 20310	2
Commander US Army Forces Command ATTN: AFOP-OM Fort McPherson, GA 30330	1
Director Program Analysis and Evaluation Office of the Secretary of Defense Room 2E330 The Pentagon Washington, DC 20310	1

Addressee	No of copies
Organization of the Joint Chiefs of Staff ATTN: J-8 Room 1D936 The Pentagon Washington, DC 20310	1
Joint Deployment Agency Deploying Systems Division MacDill Air Force Base, FL 33608	1
Commandant US Army War College Carlisle Barracks, PA 17013	2
Commandant US Army War College ATTN: Library Carlisle Barracks, PA 17013	2
Commandant US Air War College Maxwell Air Force Base, AL 36112-5532	2
Commandant US Navy War College Newport, RI 02840	1
President National Defense University ATTN: NDU-LD-CDC Washington, DC 20319-6000	1
Commandant Armed Forces Staff College Norfolk, VA 23511	1

Addressee	No of copies
Commandant US Army Command and General Staff College Fort Leavenworth, KS 66027	1
Commandant US Army Command and General Staff College ATTN: Department of Combat Development, Force Development Fort Leavenworth, KS 66027	1
Superintendent United States Military Academy ATTN: Mathematics Department West Point, NY 10996	1
Superintendent United States Military Academy ATTN: Engineering Department West Point, NY 10996	1
Superintendent Naval Postgraduate School Monterey, CA 93940	1
Naval Postgraduate School ATTN: Department of Operations Research Code 55PY (Professor Parry) Monterey, CA 93940	2
Commandant US Army Infantry School ATTN: ATSH-IVT Fort Benning, GA 31905	1
Commandant US Army Armor School ATTN: ATSB-CD Fort Knox, KY 40121-5215	1

Addressee	No of copies
Commandant US Army Field Artillery School Fort Sill, OK 73503	1
Commandant US Army Aviation School Fort Rucker, AL 36360	1
Commandant US Army Intelligence Center and School ATTN: ATSI-TD Fort Huachuca, AZ 85613	1
Commander in Chief United States Readiness Command ATTN: RCDA MacDill Air Force Base, FL 33608	1
Commander US Army Armament, Munition, and Chemical Command ATTN: DRSAR-CPB-O Rock Island, IL 61201	1
Commander US Army Combat Developments Experimentation Command Fort Ord, CA 93941	2
CINCPAC Staff C3S ASII Box 32A Camp Smith, HI 96861	2
Commander Eighth US Army APO San Francisco 96301	1

Addressee	No of copies
Commander US Army, Japan ATTN: AJCS APO San Francisco 96343	1
Assistant Chief of Staff for Intelligence Headquarters, Department of the Army Washington, DC 20310	1
Commander US Army Intelligence and Security Command ATTN: IACS Arlington Hall Station, VA 22212	1
Commander US Army Intelligence and Threat Analysis Center Arlington Hall Station, VA 22212	1
Commander/Director US Army Engineer Studies Center Casey Building, No. 2594 Fort Belvoir, VA 22060	1
Commander US Army Missile Command Redstone Arsenal, AL 35898-5090	1
Commander in Chief US Army, Europe & Seventh Army ATTN: AEAGF APO New York 09403	1
Commander in Chief US Army, Europe & Seventh Army ATTN: AEAGX-OR (Dr. Leake) APO New York 09403	1

Addressee	No of copies
Commander US Army Training and Doctrine Command ATTN: ATCD-AU Fort Monroe, VA 23651	1
Commander US Army Training and Doctrine Command Fort Monroe, VA 23651	1
Commander US Army Materiel Command 5001 Eisenhower Avenue Alexandria, VA 22333	1
Commander US Army Tank-Automotive Command Warren, MI 48090	1
Commander US Army Information Systems Command Fort Huachuca, AZ 85613	1
US Army CE Command Program Analysis and Evaluation Systems Analysis Division Fort Monmouth, NJ 07703	1
Air Force Center for Studies and Analyses AFCSA/SAMI Room 1D363, Pentagon Washington, DC 20330-5425	1
Headquarters, US Air Force Office of Worldwide Management of Studies & Analyses ATTN: AF/SAL The Pentagon Washington, DC 20330	1

Addressee	No of copies
Headquarters Foreign Technology Division (AFSC) Wright-Patterson Air Force Base, OH 45433	1
AFTEC/OAN ATTN: Modelling Branch Kirtland Air Force Base, NM 87117	1
Commander Tactical Air Command TAC Analysis Group Langley Air Force Base, VA 23665	1
Joint Studies Group (TAC) ATTN: OSS Langley Air Force Base, VA 23665	1
Commander USAF Systems Command Andrews Air Force Base Washington, DC 20334	1
Commandant Air Force Institute of Technology ATTN: AFIT-EN Wright-Patterson Air Force Base, OH 45433	1
Headquarters Rome Air Development Center (AFSC) ATTN: COAM (Mr. T. Humiston) Griffiss Air Force Base, NY 13441	1
Center for Wargaming Naval War College ATTN: Naval Warfare Gaming System Code 33 Newport, RI 02840	1

Addressee	No of copies
President Center for Naval Analyses 4401 Ford Avenue Post Office Box 16268 Alexandria, VA 22302-0268	2
Chief of Naval Operations ATTN: OP-96 Room 4A526, Pentagon Washington, DC 20350	1
Commander Naval Air Systems Command ATTN: Mission and Effectiveness Analysis Division (AIR-5264) Washington, DC 20361	1
US Army Liaison Officer Naval Weapons Center China Lake, CA 93555	1
US LNO Supreme Allied Commander Atlantic Norfolk, VA 23511-5100	1
Marine Corps Operations Analysis Group Center for Naval Analyses 4401 Ford Avenue P. O. Box 11280 Alexandria, VA 22302-0268	1
Director USMC Development and Education Center Quantico, VA 22134	1
CINC, US REDCOM ATTN: RCSTA MacDill Air Force Base, FL 33608-6001	1

Addressee	No of copies
Dr. Bruce Bennett The RAND Corporation 2100 M St., NW Washington, DC 20037-1270	1
Dr. Steven Dreshner Director, Arroyo Center 1700 Main Street Santa Monica, CA 90406	1
COL Trevor N. Dupuy Historical Evaluation and Research Organization Suite 400 8316 Arlington Boulevard Fairfax, VA 22031	1
Dr. Joseph Engle Arthur D. Little, Technology Resource Center 600 Maryland Avenue, SW Washington, DC 20002	1
Dr. Phillip Gould Manager, Forces Planning Program Institute for Defense Analyses 1801 North Beauregard Street Alexandria, VA 22311	1
Prof. Reiner K. Huber Universtat der Budeswehr Munchen Fakulatat fur Informatik Institute fur Angewandte Systemforschungs and Operations Research Werner-Heisenberg Weg 39 D-8014 Neubiberg Federal Republic of Germany	1

Addressee	No of copies
Prof Wayne P. Hughes, Jr. Code 55HL Department of Operations Research US Naval Postgraduate School Monterey, CA 93943	1
Mr. Edward Kerlin Institute for Defense Analyses 1801 North Beauregard St. Alexandria, VA 22311	1
Dr. Wilbur Payne Commandant USA Air Defense School (SG-FAADS) Ft. Bliss, Texas 79916-7050	1
Prof. Ronald W. Shephard c/o Royal Military College of Science Shrivenham Swindon, Wilts SN6 8LA England, United Kingdom	1
Prof. J. David Singer Department of Political Science University of Michigan Ann Arbor, MI 48109	1
BG William A. Stofft Director, Center of Military History US Army Casimir Pulaski Building 20 Massachusetts Avenue Washington, DC 20001	1
Mr. W. E. Sykes, Jr. GAO/NSAID Room 5112 US General Accounting Office 441 G St., NW Washington, DC 20548	1

Addressee	No of copies
Prof. James G. Taylor Code 55HL Department of Operations Research US Naval Postgraduate School Monterey, CA 93943	1
Mr. Clayton J. Thomas Chief Scientist ACS/Studies and Analyses Headquarters, United States Air Force Washington, DC 20330-5420	1
Mr. Leonard Wainstein Institute for Defense Analyses 1801 North Beauregard Street Alexandria, VA 22311	1
Mr. Herbert K. Weiss P. O. Box 2668 Palos Verdes Peninsula, CA 90274	1
Dr. Daniel Willard Office, Deputy Undersecretary of the Army (Operations Research) Room 2E660, The Pentagon Washington, DC 20310	1
Commander SHAPE Technical Centre ATTN: Mr. Rex Goad The Hague, Netherlands	1
Executive Director The Military Conflict Institute 2842 Ashby Avenue Berkeley, CA 94705-2341	1

Addressee	No of copies
Director Livermore National Laboratories ATTN: Mr. Donald Blumenthal Livermore, CA 94550	1
Commander Chemical Research and Development Command ATTN: Mr. John Seigh Aberdeen Proving Ground, MD 21005	1
NASA-CALTECH Jet Propulsion Laboratory ATTN: Dr. Joseph P. Fearey 4800 Oak Grove Drive Pasadena, CA 91109	1
Commander Combat Developments Experimentation Center ATTN: Mr. Marion Bryson Ft. Ord, CA 93941	1
Commander TRADOC Analysis Center Ft. Levenworth, KS 66027-5200	1
Mr. Forrest Waller LFW Associates P. O. Box 12328 Arlington, VA 22209-8328	1
Dr. John Ennf Director, Military Studies Abt Associates - Suite 500 4250 Connecticut Avenue, NW Washington, DC 20008	1
Mr. W. Robert Gerber 6060 Haverhill Ct. Springfield, VA 22152	1

Addressee	No of copies
Director Brookings Institution ATTN: Dr. Joshua Epstein 1775 Massachusetts Avenue, NW Washington, DC 20036	1
Office of Air Force History AF/CHO Building 5681 Bolling Air Force Base Washington, DC 20332	1
Director of the Naval Historical Center Building 57 Washington navy Yard Annex Washington, DC 20374	1
Director of Marine Corps History and Museums Henderson Hall Southgate Road and Orme Street Arlington, VA 22214	1
Director, Nuclear Assessment Directorate Defense Nuclear Agency Hybla Valley Federal Building 6801 Telegraph Road Alexandria, VA 22310	1
Dr. David L. I. Kirkpatrick British Embassy 3100 Massachusetts Avenue, NW Washington, DC 20008	1

Internal Distribution

Unclassified Library	2
RQ	30

GLOSSARY

1. ABBREVIATIONS, ACRONYMS, AND SHORT TERMS

CAA	US Army Concepts Analysis Agency
CDES	CHASE Data Enhancement Study
CHASE	Combat History Analysis Study Effort
COSH(A)	hyperbolic cosine of the quantity A
CORG	Combat Operations Research Group
CUNO(A)	CUNO (X) = cumulative normal distribution function of the number X, i.e., $\mathrm{CUNO}(X)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{X} \exp(-x^2/2)\,dx$
EXP(A)	exponential function of the quantity A, that is, the base of the natural system of logarithms raised to the power A
HERO	Historical Evaluation and Research Organization
km	kilometer(s)
LOG(A)	natural logarithm of the quantity A
MEAN	arithmetical average value
OLS	ordinary least squares
RESADV (a,b)	RESADV (a,b) = ADV - a - b * LOG (FR)
SD	standard deviation
SINH(A)	hyperbolic sine of the quantity A
sq km	square kilometer(s)
SQR(A)	square root of the quantity A
WLS	weighted least squares

2. TERMS UNIQUE TO THIS STUDY

A	Attacker's surviving personnel fraction, A = X/XO = 1 - FX.
AA	Attacker's Lanchesterian personnel activity parameter: the value of AA in dY/dt = -AA * X, where X and Y are the attacker's and the defender's current surviving personnel strengths.
ACHA	Attacker's adjudged mission accomplishment rating on a scale of 0 (mission not accomplished) to 10 (mission fully accomplished). Cf. HERO Table 5.
ACHD	Defender's adjudged mission accomplishment rating on a scale of 0 (mission not accomplished) to 10 (mission fully accomplished). Cf. HERO Table 5.
ADV	Lanchesterian defender's advantage parameter, ADV = LOG(MU).
AEROA	Relative air superiority achieved by the attacker. Cf. HERO Table 2.
AIRA	Attacker's adjudged relative air superiority. Cf. HERO Table 6.
ARTYA	Total number of artillery pieces for the attacker (0 if none present, -1 if unknown). Cf. HERO Table 3.
ARTYD	Total number of artillery pieces for the defender (0 if none present, -1 if unknown). Cf. HERO Table 3.
ATK	Attack, attacker
ATKWIN	Attaker wins, i.e., WINA = +1
BWS	Bodart-Willard-Schmieman data base
CAMPGN	Name of the campaign of which this battle/engagement is a part. Cf. HERO Table 1.
CARTYA	Number of the attacker's artillery pieces that were destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.
CARTYD	Number of the defender's artillery pieces that were destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.

CAVA	Number of mounted troops (cavalry, dragoons, and mounted infantry) for the attacker (0 if none, -1 if unknown). Cf. HERO Table 3.
CAVD	Number of mounted troops (cavalry, dragoons, and mounted infantry) for the defender (0 if none, -1 if unknown). Cf. HERO Table 3.
CEA	Attacker's adjudged relative advantage in combat effectiveness. Cf. HERO Table 4.
CER	Defender's personnel casualty exchange ratio, CER = CX/CY (see also FER).
CFP	Personnel casualty fraction product, CFP = FX * FY.
CFLYA	Number of the attacker's combat aircraft lost as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.
CFLYD	Number of the defender's combat aircraft lost as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.
COA	Name of the commander of the attacker's force unit that fought in the battle. Cf. HERO Table 1.
COD	Name of the commander of the defender's force unit that fought in the battle. Cf. HERO Table 1.
CTANKA	Number of the attacker's tanks and other AFVs destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.
CTANKD	Number of the defender's tanks and other AFVs destroyed, damaged, or captured as a result of enemy action (0 if none, -1 if unknown). Cf. HERO Table 3.
CX	Battle casualties to the attacker's personnel (0 if none, -1 if unknown). Cf. HERO Table 3.
CY	Battle casualties to the defender's personnel (0 if none, -1 if unknown). Cf. HERO Table 3.
D	Defender's surviving personnel fraction, D = Y/YO = 1 - FY.
DAR	Lanchesterian personnel activity ratio, DAR = DD / AA = (XO ** 2 - X ** 2) / (YO ** 2 - Y ** 2) = (FR ** 2) * (1 - A ** 2) / (1 - D ** 2)

DATE	Date on which the battle began, in the form ±YYYYMMDD, where YYYY is the year, MM is the month number, and DD is the number of the day of the month. DATE is positive for AD dates and negative for BC dates. Cf. HERO Table 1.
DD	Defender's Lanchesterian personnel activity parameter: the value of DD in dX/dt = -DD * Y, where X and Y are the attacker's and the defender's current surviving personnel strengths.
DEEPA	Attacker's adjudged relative depth advantage. Cf. HERO Table 6.
DEF	Defense, defender.
DEFWIN	Defender wins, i.e., WINA = -1
EPS	Lanchesterian bitterness parameter defined by the equation EPS = LOG((1 + MU)/(A + D * MU)).
FER	Defender's personnel fractional exchange ratio, FER = FX/FY = CER/FR (see also CER).
FLYA	Total number of air sorties flown in support of the attacker (0 if none flown, -1 if unknown). Cf. HERO Table 3.
FLYD	Total number of air sorties flown in support of the defender (0 if none flown, -1 if unknown). Cf. HERO Table 3.
FORTSA	Attacker's adjudged relative fortification advantage. Cf. HERO Table 6.
FPREPA	Attacker's adjudged relative force preponderance. Cf. HERO Table 6.
FR	Attacker's personnel force ratio, FR = XO/YO.
FX	Attacker's personnel casualty fraction, FX = CX/XO.
FY	Defender's personnel casualty fraction, FY = CY/YO.
INITA	Attacker's adjudged relative advantage in initiative. Cf. HERO Table 4.
INTELA	Attacker's adjudged relative advantage in (military) intelligence. Cf. HERO Table 4.

ISEQNO	Index or sequence number of the battle in the computerized data base (see Appendix H for an index of the computerized data base battles by ISEQNO).
KPDA	Attacker's average rate of advance, in kilometers per day. Positive values indicate an attacker's advance, negative ones a defender's advance, and zero values either no or a negligible advance. The value -9999 is used if the average rate of advance is unknown. Cf. HERO Table 5.
LAMBDA	Lanchesterian intensity parameter, LAMBDA = EPS/T.
LEADA	Attacker's adjudged relative advantage in leadership. Cf. HERO Table 4.
LEADAA	Attacker's adjudged relative leadership advantage. Cf. HERO Table 6.
LOCN	Name of the place where the battle occurred (usually a nation or other geopolitical region). Cf. HERO Table 1.
LOGSA	Attacker's adjudged relative advantage in logistics. Cf. HERO Table 4.
LOGSAA	Attacker's adjudged relative logistics advantage. Cf. HERO Table 6.
LTA	Total number of light armored tank-like vehicles for the attacker (0 if none present, -1 if unknown). Cf. HERO Table 3.
LTD	Total number of light armored tank-like vehicles for the defender (0 if none present, -1 if unknown). Cf. HERO Table 3.
MANA	Attacker's adjudged relative maneuver advantage. Cf. HERO Table 6.
MAX	Maximum.
MAX.L	Maximum likelihood value.
MBTA	Total number of main battle tanks for the attacker (0 if none present, -1 if unknown). Cf. HERO Table 3.
MBTD	Total number of main battle tanks for the defender (0 if none present, -1 if unknown). Cf. HERO Table 3.
MIN	Minimum.

MOBILA	Attacker's adjudged relative mobility superiority. Cf. HERO Table 6.
MOMNTA	Attacker's adjudged relative advantage in momentum. Cf. HERO Table 4.
MORALA	Attacker's adjudged relative advantage in morale. Cf. HERO Table 4.
MU	Lanchesterian mu-parameter, MU = SQR(DAR) / FR.
NAMA	Name of the attacker's force element that fought the battle. Cf. HERO Table 1.
NAMD	Name of the defender's force element that fought the battle. Cf. HERO Table 1.
NAME	Name of the battle or engagement. Cf. HERO Table 1.
NN	The total number of battles in the data base.
PLANA	Attacker's adjudged relative planning effectiveness. Cf. HERO Table 6.
POSTD1	Defender's primary defensive posture. Cf. HERO Table 2.
POSTD2	Defender's secondary defensive posture. Cf. HERO Table 2.
PRIA1	Attacker's primary tactical scheme, part 1. Cf. HERO Table 7.
PRIA2	Attacker's primary tactical scheme, part 2. Cf. HERO Table 7.
PRIA3	Attacker's primary tactical scheme, part 3. Cf. HERO Table 7.
PRID1	Defender's primary tactical scheme, part 1. Cf. HERO Table 7.
PRID2	Defender's primary tactical scheme, part 2. Cf. HERO Table 7.
PRID3	Defender's primary tactical scheme, part 3. Cf. HERO Table 7.
QUALA	Attacker's adjudged relative force quality. Cf. HERO Table 6.

RESA	Attacker's adjudged relative skill in use of reserves. Cf. HERO Table 6.
RESOA1	Attacker's resolution/outcome, part 1. Cf. HERO Table 7.
RESOA2	Attacker's resolution/outcome, part 2. Cf. HERO Table 7.
RESOA3	Attacker's resolution/outcome, part 3. Cf. HERO Table 7.
RESOD1	Defender's resolution/outcome, part 1. Cf. HERO Table 7.
RESOD2	Defender's resolution/outcome, part 2. Cf. HERO Table 7.
RESOD3	Defender's resolution/outcome, part 3. Cf. HERO Table 7.
SECA1	Attacker's secondary tactical scheme, part 1. Cf. HERO Table 7.
SECA2	Attacker's secondary tactical scheme, part 2. Cf. HERO Table 7.
SECA3	Attacker's secondary tactical scheme, part 3. Cf. HERO Table 7.
SECD1	Defender's secondary tactical scheme, part 1. Cf. HERO Table 7.
SECD2	Defender's secondary tactical scheme, part 2. Cf. HERO Table 7.
SECD3	Defender's secondary tactical scheme, part 3. Cf. HERO Table 7.
SKEW	Coefficient of skewness (see following paragraph 4, Definitions).
SURPA	Relative surprise achieved by the attacker. Cf. HERO Table 2.
SURPAA	Attacker's adjudged relative surprise advantage. Cf. HERO Table 6.
T	Duration of the battle, in days, an integer. Cf. HERO Table 1.

TANKA	Total number of armored tank-like vehicles for the attacker (includes tanks; armored, self-propelled tank guns; and armored assault guns) (0 if none present, -1 if unknown). Cf. HERO Table 3.
TANKD	Total number of armored tank-like vehicles for the defender (includes tanks; armored, self-propelled tank guns; and armored assault guns) (0 if none present, -1 if unknown). Cf. HERO Table 3.
TECHA	Attacker's adjudged relative advantage in technology. Cf. HERO Table 4.
TERRA	Attacker's adjudged relative terrain/roads advantage. Cf. HERO Table 6.
TERRA1	Three-character primary terrain descriptor. Cf. HERO Table 2.
TERRA2	Three-character secondary terrain descriptor. Cf. HERO Table 2.
TRNGA	Attacker's adjudged relative advantage in training and experience. Cf. HERO Table 4.
WAR	Name of the war of which the battle/engagement is a part. Cf. HERO Table 1.
WGT	Relative adjudged rating of the accuracy/validity of the data for this battle (not used in this paper).
WINA	Attacker's adjudged relative level of victory, i.e., WINA = +1 when the attacker wins, WINA = -1 when the defender wins, and WINA = 0 when the battle is a draw. Cf. HERO Table 5.
WOF	Width of front, in kilometers. Cf. HERO Table 1.
WX1	First five-character weather, season, and climate descriptor. Cf. HERO Table 2.
WX2	Second five-character weather, season, and climate descriptor. Cf. HERO Table 2.
WX3	Third five-character weather, season, and climate descriptor. Cf. HERO Table 2.
WXA	Attacker's adjudged relative weather advantage. Cf. HERO Table 6.
X	Attacker's surviving personnel strength, X = X0 - CX.

XKURT	Coefficient of excess kurtosis (see following paragraph 4, Definitions).
XO	Total engaged personnel strength of the attacker (-1 if unknown). Cf. HERO Table 3.
Y	Defender's surviving personnel strength, Y = YO - CY.
YO	Total engaged personnel strength of the defender (-1 if unknown). Cf. HERO Table 3.

3. MODELS, ROUTINES, AND SIMULATIONS

BINMAKER	Prepares histograms and contingency tables.
DALOFIT	Performs logistic regression by fitting multivariate logistic functions using the maximum likelihood method (see logistic regression in following paragraph 4, Definitions).
DATAMAKER	Reads the computerized HERO data base and prepares data files for other programs.
ROSEPACK	Finds robust multivariate regression fits to data.
SPSS	Statistical Package for the Social Studies.
UNIVARIATE	Finds empirical distribution functions and compares them to theoretical distribution functions.

4. DEFINITIONS

adjusted advantage
Empirically estimated value of the ADV parameter, calculated after adjusting strengths for presumed reinforcements and replacements as explained in paragraph 4-3b(4).

advantage
Synonym for defender's advantage or for ADV, q.v.

bitterness
Synonym for EPS, q.v.

BWS data base
Bodart-Willard-Schmieman data base **(Ref 2-5)**. This data base originated with Bodart's Kriesslexicon **(Ref 2-6)**, which was originally computerized by Willard **(Ref 2-7)**, and later modified by Schmeiman **(Ref 2-8)**.

coefficient of excess kurtosis

Symbolized by XKURT, and defined by the formula

$$XKURT = m4 / (SD)^4 - 3,$$

where SD is the standard deviation and m4 is the fourth-order moment about the mean, that is,

$$m4 = (n^{-1}) * SUM (for\ i = 1\ to\ n\ of\ (x_i - MEAN)^4)$$

where MEAN is the mean of the x_i values. XKURT is zero for the normal distribution. XKURT tends to be positive for distributions that are "fatter-tailed," and negative for those that are "thinner-tailed," than the normal frequency function. The SD of XKURT is approximately equal to SQR(24/n), where n is the sample size **(Refs G-1 and G-2).**

coefficient of skewness

Symbolized by SKEW and defined by the formula

$$SKEW = m3/(SD)^3,$$

where SD is the standard deviation and m3 is the third-order moment about the mean, that is

$$m3 = (n^{-1}) * SUM (for\ i = 1\ to\ n\ of\ (x_i - MEAN)^3)$$

where MEAN is the mean of the x_i values. SKEW is zero for any distribution of values symmetric about their mean value--in particular it is zero for the normal distribution. SKEW tends to be positive for distributions with a "long tail" above the mean, and negative for distributions with a "long tail" below the mean. The standard deviation of SKEW is approximately equal to SQR(6/n), where n is the sample size **(Refs G-1 and G-2).**

computerized data base

The computerized version (prepared by CAA in late 1984 and early 1985) of the tabular data in the HERO data base, and described in Appendices F through H of this paper.

CORG data base

Data base complied by the Combat Operations Research Group (CORG) in the early 1960s **(Refs 2-2 through 2-4).**

empirical distribution

Function whose value at x is defined to be the fraction of data items with values less than x.

exploratory subsample

A sample of 100 battles selected randomly from those computerized data base battles whose starting dates are earlier than 1 January 1943.

factor analysis
A statistical technique for reducing the level of redundancy in the data.

force ratio
Synonym for attacker's force ratio or FR, q.v.

intensity
Synonym for LAMBDA, q.v.

HERO data base
The data base prepared for the US Army Concepts Analysis Agency (CAA) by HERO under Contract No. MDA903-82-C-0363, published by CAA in September 1984 as "Analysis of Factors That Have Influenced Outcomes of Battles and Engagements," CAA-SR-84-6, in six volumes as follows:

Vol.	DTIC No.	Title
I	B086 797L	Main Report
II	B087 718L	HERO Summary and Introductory Materials; Part One: Wars of the 17th, 18th, and 19th Centuries; Vol. II: Wars from 1600 through 1800.
III	B087 719L	Part One: Wars of the 17th, 18th, and 19th Centuries; Vol. III: Wars from 1805 through 1900.
IV	B087 720L	Part Two: Wars of the 20th Century; Vol. IV: Wars from 1904-1940.
V	B087 721L	Part Two: Wars of the 20th Century; Vol. V: World War II, 1939-1945; Campaigns in North Africa, Italy, and Western Europe.
VI	B087 722L	Part Two: Wars of the 20th Century; Vol. VI: World War II, 1939-1945; Campaigns in France, 1940, on the Eastern Front, and of the War Against Japan. The 1967, 1968, and 1973 Arab-Israeli Wars.

logarithmic
Natural logarithm of, as in "The logarithmic force ratio is a synonym for LOG(FR)."

logistic regression
A statisical technique for fitting a logistic function to the probability of responses to an administered dose or other stimulus. Here the responses are treated as categorical (discrete), for example, as either a win, a loss, or a draw (see Appendix J for a discussion of logistic regression).

Ockham's Razor

The epistemological principle that "Entities are not to be multiplied without necessity." That is, the fewest assumptions and the simplest formulae are to be used unless the data can be explained only through the use of additional factors or mathematical complexity.

Prob. Kolmog. exceedance

The probability that the Kolmogoroff test criterion is exceeded. That is, the probability that the absolute deviation between a theoretical and an empirical distribution function would be exceeded by chance, even though the empirical distribution function is for a random sample from that theoretical distribution function.

residual advantage

Synonym for RESADV, q.v.

sample size

Number of data points used

CAA — Concepts Analysis Agency, United States Army	**COMBAT HISTORY ANALYSIS STUDY EFFORT (CHASE): PROGRESS REPORT FOR THE PERIOD AUGUST 1984 - JUNE 1985**	**STUDY SUMMARY CAA-TP-86-2**

<u>THE REASON FOR PERFORMING THE STUDY</u> was to carry out the initial phase of the Combat History Analysis Study Effort (CHASE), whose ultimate purpose is to search for historically-based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses.

<u>THE PRINCIPAL FINDING</u> of the work done during the period covered by this paper (August 1984 to June 1985) is that data on historical battles can be used to discover quantitative trends and relations of potential significance to military operations research, concept formulation, wargaming, and studies and analyses.

<u>THE MAIN ASSUMPTIONS</u> on which the CHASE Study, as well as its major phases, rests are:

(1) Historical battle data can be analyzed using modern statistical methods.

(2) Formulas are not to be complicated without good empirical evidence.

(3) Long-term trends and relations can be extrapolated to future situations with a reasonable degree of confidence.

<u>THE PRINCIPAL LIMITATIONS</u> which may affect the findings presented in this progress report are as follows:

(1) Data on strengths at intermediate stages during the course of a battle were not available for use in this phase of the CHASE Study.

(2) The study used a data base prepared for the US Army Concepts Analysis Agency (CAA) by the Historical and Research Evaluation Organization (HERO). The HERO data base, even though composed of 601 battles, is still not large enough to support adequately all of the statistical analyses that should be performed.

(3) Typographical mistakes, omissions, ambiguities and ill-defined data categories in the HERO data base weakened some of the analysis results, and precluded some analyses that would have been desirable.

(4) Because of data inadequacies and the limited scope of this initial phase of the CHASE Study, not all of CHASE's Essential Elements of Analysis (EEAs) could be fully addressed.

THE SCOPE OF THE WORK done during the period covered by this progress report, was limited to an initial analysis of the HERO data base of 601 battles. This scope included:

(1) Reducing to machine-readable form all of the tabulated data in the HERO data base.

(2) Assessing the suitability of the data base for quantitative analysis.

(3) Summarizing selected portions of these data to facilitate their efficient use in military operations research, concept formulation, wargaming, and studies and analyses.

(4) Seeking important trends and interrelations present but hidden in these data.

(5) Testing selected hypotheses against the data.

THE STUDY OBJECTIVE for the period covered by this progress report included:

(1) Evaluating the suitability of the HERO data base for quantitative analysis, identifying essential data base improvements, and taking necessary corrective measures.

(2) Experimenting with a variety of analytical techniques to assess their ability to expose quantitative trends and relations of significant potential use in military operations reserch, concept formulation, wargaming, and studies and analyses.

(3) Identifying specific issues for further investigation in subsequent phases of the CHASE Study.

THE STUDY SPONSOR was the US Army Concepts Analysis Agency.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Resources and Requirements Directorate.

COMMENTS AND SUGGESTIONS may be sent to the Director, US Army Concepts Analysis Agency, ATTN: CSCA-RQ, 8120 Woodmont Avenue, Bethesda, MD 20814-2797.